

DigitalHouse >

DATA SCIENCE

Programa do curso

Quem somos

Somos um **centro de formação de habilidades digitais** que desafia a repensar tudo, **aprendendo a passar da ideia para a ação**. Incorporamos o conceito de Coding School, oferecendo cursos intensivos de 5 meses sobre desenvolvimento de aplicativos web, mobile, marketing digital, UX Design, data science e analytics, entre outros.

Sabemos que a melhor maneira de ensinar estas disciplinas é com uma metodologia de “aprender fazendo” que **integre, permanentemente, a parte teórica a casos práticos**.

Queremos que os alunos tragam ideias e possam desenvolvê-las, recebendo um feedback constante dos professores e instrutores, que, por serem especialistas nesse setor atual, poderão orientá-los no caminho da aprendizagem.

Queremos passar do uso da tecnologia para sermos protagonistas da sua criação. Por isso, na Digital House formamos as novas gerações de programadores e profissionais do mundo digital, para que eles sonhem e inovem com sucesso.

Buscamos treinar e estimular os nossos alunos para que rapidamente possam fazer parte de equipes de empresas globais, criar suas startups ou trabalhar como freelancers em qualquer lugar do mundo.

Para quem é o curso?



Perfil do aluno

ESTUDANTES, PROFISSIONAIS DE DIVERSAS ÁREAS

Economistas, administradores de empresas, cientistas sociais, físicos, químicos, médicos, entre outros que desejam aprender Data Science (Ciência de Dados) para conquistar novas habilidades, criar novos projetos e melhorar sua comunicação com engenheiros de Big Data ou especialistas em Machine Learning.

PROFISSIONAIS DE MARKETING

que queiram incorporar ferramentas de análise de dados para inovar nas suas estratégias de segmentação, personalização da oferta e previsões de todos os tipos sobre os clientes.

ENGENHEIROS DE SISTEMAS E ANALISTAS DE SISTEMAS

que queiram se especializar, incorporar os conhecimentos de estatística e Machine Learning, e adquirir habilidades práticas com as ferramentas específicas para fazer análises e organização de dados.

EMPREENDEDORES

que queiram criar seu próprio negócio com base em dados e/ou técnicas de inteligência artificial.

PROGRAMADORES

que precisam se atualizar e conhecer as principais técnicas da ciência de dados.

CIENTISTAS

de qualquer área em busca de ferramentas mais sofisticadas para realizar suas análises de dados.

Programa do curso

Ao final do curso, os participantes serão capazes de:

- » Extrair, consultar, limpar e agregar dados para sua análise.
- » Realizar análises visuais e estatísticas de dados, usando o Python e suas bibliotecas associadas.

- » Construir, implementar e avaliar problemas de Data Science usando os algoritmos apropriados de Machine Learning.
- » Usar as ferramentas de visualização adequadas para comunicar suas conclusões.
- » Criar relatórios claros para as diversas partes envolvidas.
- » Investigar, modelar e validar processos de resolução de problemas aplicados a datasets provenientes de diversos setores para oferecer experiências em vários tipos de problemas e soluções do mundo real.

Proposta do curso



- » A necessidade de contar com analistas de dados qualificados cresce rapidamente em todos os setores empresariais e industriais, assim como nos âmbitos onde se desenvolvem novos paradigmas educacionais, sociais e científicos.
- » O curso de Data Science da DH oferece conhecimentos sólidos no processo de extração, processamento e análise de grandes quantidades de dados, com o objetivo de apoiar a tomada de decisões estratégicas em organizações de todos os tipos.
- » O programa está estruturado com base no ciclo de vida completo dos dados, da sua coleta e preparação até a sua análise e visualização efetiva.
- » Você aprenderá as habilidades e os conhecimentos fundamentais para ter o seu primeiro trabalho como analista de dados.
- » E descobrirá os caminhos possíveis para continuar a sua aprendizagem nos temas de especialização que sejam do seu interesse.

Programa do curso

01. INTRODUÇÃO A DATA SCIENCE

MÓDULO 1: Fundamentos: Numpy, Stats, SQL e Visualização

***Python e Numpy:** Demonstrar conceitos de programação usando as ferramentas Python e Numpy para navegar fontes de dados e coleções.

***Estatística Descritiva:** Rever e aplicar os fundamentos de estatística descritiva.

***SQL e Bancos de Dados:** Conhecer diferentes tipos de bancos de dados, rever expressões SQL e realizar atividades práticas obtendo dados de um banco remoto.

***Introdução a Gráficos e Visualização:** Realizar atividades práticas usando notebooks iPython e plot.ly para visualizar dados.

DESAFIO 1

Com um dataset de pontuações SAT dos EUA, os participantes realizarão uma análise exploratória utilizando Numpy e Matplotlib, aplicando técnicas básicas de estatística descritiva.

MÓDULO 2: Exploratory Data Analysis (EDA), Pandas e SciPy

***Pandas e Tabelas Pivot:** Introdução a Pandas (biblioteca para ler, limpar, realizar o parsing e representar por gráficos dados usando funções booleanas, indexação, séries, joins e outras funcionalidades).

***Limpeza de dados:** Conceitos de “tidy data”, estruturas de dados adequadas, introdução ao problema de dados omissos, noções de Expressões Regulares.

***Estatística Inferencial:** Noções de probabilidade, distribuições amostrais, intervalos de confiança, testes de hipótese. Uso de biblioteca SciPy.

DESAFIO 2

Usando um dataset CRU de hits da “Billboard”, os participantes utilizarão Pandas para limpar os dados. Depois apresentarão formalmente um problema e realizarão uma análise exploratória para um editor de música.

02. INTRODUÇÃO A MACHINE LEARNING, REGRESSÃO, AVALIAÇÃO DE MODELOS, WEB SCRAPING

MÓDULO 3: Introdução a machine learning. Regressão Linear, StatsModels e Scikit-Learn

***Machine Learning:** Introdução a conceitos nodais: desvio, variância, overfitting, underfitting. Identificação de diferentes tipos de algoritmos de Machine Learning (supervisionados e não supervisionados). Formas de estimar o erro de generalização (trainteste split, crossvalidation). Apresentação geral das técnicas mais utilizadas.

Regressão Linear Simples e Múltipla: Interpretação, estimativa de coeficientes beta, suposições, medidas de ajuste. Introdução de não linearidades no modelo.

Statsmodels e Sklearn: Introdução ao uso das duas bibliotecas para estimar modelos estatísticos e algoritmos de Machine Learning na Python. API, tipos e objetos na ScikitLearn.

Regularização e Otimização: Noções gerais de regularização. Regressão LASSO e Ridge. Feature scaling.

DESAFIO 3

A partir do acesso a determinados dados de vendas, os participantes poderão escolher entre realizar uma investigação de marketing ou uma auditoria tributária, usando Pandas, Statsmodels e Sklearn para transformar os dados, fazer uma regressão linear e visualizar os resultados.

MÓDULO 4: Problemas de Classificação, Tuning de Hiperparâmetros e Web Scraping

Introdução aos Problemas de Classificação: Problema dos KNearest Neighbours.

Algoritmos usados para problemas de classificação: Regressão Logística. Naïve Bayes, Support Vector Machines. Avaliação de algoritmos de classificação: métricas de erro, acurácia, limitações de acurácia. Precisão, recall, F1Score, Curva ROC, área sob a curva (AUC).

Ajuste de hiperparâmetros: Definição de hiperparâmetro. Métodos de busca: busca exaustiva (grid search), busca aleatorizada (random search), busca por algoritmos genéticos. Dimensionalidade.

Web Scraping: Noções de HTML (Palavraschave, árvores). Uso de bibliotecas urllib2 e BeautifulSoup para scrapear sites.

DESAFIO 4

Colocando-se na posição de fornecedores do governo, os participantes vão fazer scraping de um site e utilizar Pandas, Statsmodels e NLTK para analisar dados, fazer regressão logística e avaliar coeficientes de correlação.

03. APIS E MACHINE LEARNING AVANÇADO: ANÁLISE DE SÉRIES DE TEMPO, MÉTODOS DE ENSEMBLE

MÓDULO 5: Séries de tempo, Classificação e Pipelines

***Séries de Tempo:** Pré-processamento de séries. Enfoque clássico (tendência, ciclo e resíduo). Sazonalidade. Modelos básicos: média constante, tendência determinística, média móvel, suavização exponencial simples. Modelos avançados: processos AR, MA, ARMA, ARIMA e ARIMAX.

Pipelines e Custom Transformers em ScikitLearn: Pipelines: automatização e encapsulamento de etapas do workflow. Custom Transformers: noção de objeto e tipo em Python.

Seleção de variáveis: Motivação de seleção de variáveis. Filter Methods, Wrapper Methods, Embedding Methods.

TRABALHO INTEGRADOR - ETAPA 1:

Os alunos deverão fazer uma apresentação relâmpago para a turma, expor a definição do problema e a seleção do dataset(s) adequado(s) para abordá-lo.

MÓDULO 6: APIS, Árvores e Ensemble Learning

JSON e APIs: Estrutura de Dados JSON. Paralelismo com dicts na Python. Uso de uma API.

Árvores de Decisão: Construção de árvores por meio do algoritmo CART. Evitando o overfitting e underfitting em árvores.

Modelos de Ensemble: Noção de Ensemble Learning. Diferenças e vantagens. Apresentação do meta-algoritmo Bagging e de duas aplicações para árvores de decisão: Random Forest e ExtraTrees. Apresentação do meta-algoritmo de Boosting e de AdaBoost e Gradient Boosting. Entrando na caixa preta: feature importance e partial dependence plot em modelos de ensemble.

TRABALHO INTEGRADOR - ETAPA 2

Os alunos deverão apresentar um relatório de resultados preliminares (de carácter técnico) e fazer uma apresentação não técnica na qual exponham: a) a definição do problema; b) uma análise exploratória do dataset selecionado; e c) uma exploração preliminar de técnicas de modelagem.

04. ALGORITMOS NO SUPERVISADOS

MÓDULO 7: Clustering e Redução de Dimensionalidade

Clustering: Algoritmo K-Means, clustering hierárquico, métricas para avaliação de algoritmos de clustering.

Redução de Dimensionalidade: Introdução ao PCA. Apresentação de Análise de Correspondência Múltipla. Manifold Learning: Técnicas para redução de dimensionalidade não lineares: Escala Multidimensional (MDS), Locally Linear Embeddings e IsoMaps. Diferenças em relação ao PCA. Métricas de distância utilizadas.

TRABALHO INTEGRADOR - ETAPA 3:

Apresentação do Trabalho Integrador. Os alunos deverão entregar um notebook iPython com código, visualizações e notas técnicas para os colegas do curso. Também deverão expor as conclusões, com terminologia não técnica, para um público leigo.

Requisitos Mínimos

Os candidatos a alunos devem mostrar conhecimento sólido de estatística descritiva (nível de medição de variáveis, medidas de tendência central e dispersão, etc.) e noções de programação (estruturas de dados e de controle, definição de funções, etc.). É recomendável (mas não obrigatório) ter noções de estatística inferencial (amostragem, intervalos de confiança, teste de hipótese e distribuições de probabilidade), além de já ter tido contato com a programação em Python.



Nosso campus

Queremos ser uma comunidade que vá além da sala de aula, uma comunidade onde surjam os próximos projetos que transformarão os mais diversos setores.



Processo de admissão

A disponibilidade de vagas nos cursos é limitada e você deverá passar pelo seguinte processo de admissão:

01

PARA SE CANDIDATAR

Solicitar e preencher o formulário de inscrição no site

FORMULÁRIO DE INSCRIÇÃO

02

ENTREVISTA DE ADMISSÃO

Agende uma entrevista para conhecermos o seu perfil, seus interesses e expectativas com relação ao curso.

03

DESAFIO ON-LINE

Faça um exercício real que enviaremos para podermos entender a sua forma de pensar e de resolver situações.

FAQs

POSSO FAZER O CURSO SEM TER EXPERIÊNCIA E/OU CONHECIMENTOS PRÉVIOS?

Sim. Na Digital House você aprenderá a analisar dados do zero. Enviaremos um conteúdo online para você se preparar antes do início das aulas. Assim você aproveitará os 5 meses ao máximo.

AS VAGAS SÃO LIMITADAS?

Sim. Todos os cursos têm no máximo 40 alunos.

COMO É A ENTREVISTA DO PROCESSO DE ADMISSÃO?

A entrevista tem como objetivo conhecer o seu interesse por aprender a ciência dos dados, os seus motivos e objetivos. Se considerarmos que o curso coincide com as suas expectativas, avaliaremos a sua capacidade de resolver problemas e situações complexas, por meio de alguns exercícios de lógica e de perguntas técnicas básicas.

PRECISO TER CONHECIMENTOS MATEMÁTICOS?

Embora em alguns cursos de tecnologia não seja rigorosamente necessário ter conhecimentos de matemática, no caso das ciências de dados precisamos que você domine alguns conceitos de cálculo e álgebra (de nível de ensino médio).

SE JÁ TENHO CONHECIMENTOS BÁSICOS DE PROGRAMAÇÃO E ESTATÍSTICA VALE A PENA FAZER O CURSO?

Sim. No nosso curso você integrará metodologias e ferramentas para os seus projetos, orientado pelos melhores especialistas do setor e do universo acadêmico. Além disso, terá como colegas pessoas de setores variados, o que agrega um enorme valor à experiência e permite estabelecer relacionamentos de trabalho para o futuro.

PRECISO COMPRAR OU LEVAR UM NOTEBOOK PARA FAZER O CURSO?

Você não precisa trazer o seu computador, já que as salas de aula oferecem os equipamentos necessários – mas, se você se sentir mais à vontade usando o seu notebook, pode trazê-lo. De qualquer forma, é recomendável que você conte com um notebook ou desktop em casa para praticar.

QUE TIPO DE COMPROMISSO IMPLICA ESTUDAR NA DIGITAL HOUSE?

Na Digital House você terá 10,5 horas de aula por semana durante 5 meses, o que representa um total de 210 horas. Mas não termina aí. Para aproveitar o curso ao máximo, você terá que (e vai querer!) praticar fora da aula, fazer exercícios e trabalhar nos projetos que forem pedidos.

HÁ ALGUMA PROVA PARA ENTRAR? EM QUE CONSISTE? É OBRIGATÓRIA?

Alguns conhecimentos básicos você pode aprender online e por conta própria, com a ajuda do nosso Material de Nivelamento. Depois você fará uma prova ou um teste que chamamos de Desafio Online para avaliar seus conhecimentos e habilidades básicas como programação em Python e conceitos de estatística.

QUEM SERÃO OS MEUS PROFESSORES?

Um dos nossos principais ativos são os professores. Eles são profissionais da área, reconhecidos e bem-sucedidos, motivados e com vontade de ensinar usando estes novos modelos de educação. Além disso, são capacitados internamente pelo nosso departamento acadêmico e pedagógico para que os conteúdos estejam sempre alinhados. Algumas aulas contam ainda com a participação de especialistas que são referência no setor.

POSSO LEVAR CONVIDADOS?

Você poderá trazer amigos e colegas à sede da Digital House, mas eles só poderão ficar com você nos espaços comuns e de co-learning, não nas salas de aula.

**Se você tiver alguma dúvida, entre em contato conosco.
Estamos aqui para te ajudar!**



Quer saber mais sobre o curso e como ele vai ajudar a transformar sua carreira?

Venha para o **Papo Digital**, um evento gratuito que acontece semanalmente aqui na **Digital House**, onde você pode **conhecer nosso campus** e conversar com os professores para tirar todas as suas dúvidas.

RESERVE A SUA VAGA



DigitalHouse >



AV. DR. CARDOSO DE MELO, 90

VILA OLÍMPIA | SÃO PAULO - SP

HORÁRIO ADMINISTRATIVO

DAS 09H00 ÀS 18H00

CONTATO@BR.DIGITALHOUSE.COM

BR.DIGITALHOUSE.COM

TEL.: (11) 4858-4000

DigitalHouse >

AV. DR. CARDOSO DE MELO, 90

VILA OLÍMPIA | SÃO PAULO - SP

HORÁRIO ADMINISTRATIVO

DAS 09H00 ÀS 18H00

CONTATO@BR.DIGITALHOUSE.COM

BR.DIGITALHOUSE.COM

TEL.: (11) 4858-4000