Introdução à Ciência de Dados
2.0
Equipe Data Science Academy

Esse e-book não pode ser usado para fins comerciais, mas pode ser distribuído livremente sob a licença Creative Commons.

Pedimos apenas, a gentileza de citar a fonte, pois todo este material é resultado de trabalho árduo de nossa equipe.

Esta foi a forma que encontramos de contribuir com a sociedade que deve ter como prioridade.

Data Science Academy

O futuro é aqui.

Introdução à Ciência de Dados 2.0

Prezado(a) Aluno(a),

Seja muito bem-vindo(a) ao curso:

Introdução à Ciência de Dados 2.0

Data Science Academy

A Data Science Academy (DSA) é um portal de ensino online especializado em Big Data, Machine Learning, Inteligência Artificial, Desenvolvimento de Chatbots e tecnologias relacionadas. Nosso objetivo é fornecer aos alunos conteúdo de alto nível por meio do uso de computador, tablet ou smartphone, em qualquer lugar, a qualquer hora, 100% online e 100% em português.

Nossa Escola

Data Science Academy

Data Science Academy - Localização

No Brasil e no Mundo.

Treinamentos Gratuitos DSA

Acreditamos que aprender não para nunca.

Big Data Fundamentos 2.0

Introdução à Ciência de Dados 2.0

Python Fundamentos Para Análise de Dados

Microsoft Power BI Para Data Science

Treinamentos preparados especialmente para Você!

Curso Introdução à Ciência de Dados 2.0

Conteúdo do Curso

1. Introdução
2. O que é Ciência de Dados?
3. Ciência de Dados e Big Data
4. Ciência de Dados e Estatística
5. Aprendizado de Máquina (Machine Learning)

Curso Introdução à Ciência de Dados 2.0

Conteúdo do Curso

6 Aplicações da Ciência de Dados

7 Ciclo de Vida de Projetos em Data Science

8 Carreiras em Data Science

9 Como se Tornar um Cientista de Dados

10 Avaliação Final e Certificado de Conclusão

Curso Introdução à Ciência de Dados 2.0

Este não é um curso técnico. Temos muitos outros cursos abordando em detalhes todas as tecnologias que envolvem a Ciência de Dados. Nosso objetivo aqui será definir claramente os conceitos e orientá-lo para o mercado de trabalho.

Aliás, já fizemos um trabalho de análise de dados para você! Coletamos, filtramos, limpamos, analisamos e estamos trazendo insights preciosos sobre Data Science. Espero que você goste!

Pré-requisitos

Muita Vontade de Aprender!

Desejo de entrar no Mundo da Ciência de Dados, uma das áreas mais promissoras para os próximos anos!

Público Alvo

- Profissionais que buscam uma carreira em Data Science e Big Data
- Analistas de Dados
- Cientistas de Dados
- Engenheiros de Dados
- Analistas de Negócios
- Administradores de Sistemas
- Desenvolvedores
- Arquitetos de Soluções
- Profissionais de TI
- Gestores de TI
- Estudantes
- Lifetime learners

O Que Esperamos de Você!

Sua Abordagem no curso.

1 a 2 horas

de dedicação por semana.

Divirta-se

Comunique-se, aprenda e divirta-se em nossa Comunidade.

Leitura do Material

E-books e material complementar!

Bibliografia

Leia a bibliografia adicional, acesse os links úteis e realize os quizzes ao final dos capítulos.

Interação

Utilize nossas Apps e interaja na rede com outros alunos, no fórum exclusivo e na timeline da Comunidade em nosso site.

Objetivos ao concluir o Curso Introdução à Ciência de Dados 2.0

Avaliação Final

Questões	Tentativas	Aproveitamento
50	3	70%

Certificado de Conclusão

E-book

Seja aprovado,

E receba o e-book com todo o conteúdo do curso de forma gratuita.

Suporte

Em todos os nossos cursos, gratuitos e pagos, o aluno recebe suporte em até 24 horas, incluindo finais de semana e feriados.

Utilize um dos nossos canais de comunicação e obtenha suporte sempre que precisar!

Nossa equipe é obcecada pelo sucesso dos nossos alunos!

Oferecemos mais do que cursos online. Oferecemos uma experiência de aprendizagem!

Compartilhe seu Certificado de Conclusão

Apps Gratuitas para iOS e Android

Data Science Academy

O que é Ciência de Dados?

Definição de Ciência de Dados

Enquanto nossas vidas continuam migrando para a internet, produzimos um fluxo constante e exaustivo de informação digital.

Estima-se que 90% dos dados armazenados no mundo foram produzidos apenas nos últimos dois anos e os rastros desses dados continuam duplicando a cada ano.

O termo “Ciência de Dados” ou Data Science em inglês, tem sido cada vez mais usado e por uma boa razão.

É uma das áreas com maior crescimento atualmente!

Um Mercado em Expansão

E a profissão de Cientista de Dados vem ganhando espaço no mercado com uma demanda cada vez maior em todo mundo!

Um Mercado em Expansão

Dentre as áreas de atuação que vem demandando esses especialistas, destacam-se o varejo, saúde, finanças, telecomunicações, segurança, robótica, reconhecimento de voz, economia digital, ciências biológicas e médicas e muitas outras.

Um Mercado em Expansão

Em pesquisa divulgada pela Computerworld, 25% das empresas afirmam planejar a contratação de profissionais de Big Data e Data Science neste ano – gerando nada menos que 3,6 milhões de empregos em todo mundo. Segundo previsão do Gartner, 400 mil vagas para profissionais de Big Data e Data Science serão abertas no Brasil em 2018.

Cientista de Dados

O professional mais procurado atualmente do Mercado.

Data Science Academy

Análises avançadas de dados serão usadas por mais de 50% das grandes empresas a partir de 2018.

- Fontes: Computer word, IDG e Catho

Data Science Academy

“Decisões baseadas em emoções não são decisões. São instintos.”

Data Science Academy

Data Science Academy

O que é Ciência de Dados?

Mas o que é Ciência de Dados afinal?

O que é Ciência de Dados?

Ciência de Dados é o termo usado para definir a extração de insights de dados que são coletados de várias fontes.

O que é Ciência de Dados?

Utilizando várias técnicas, incluindo modelagem preditiva, a Ciência de Dados ajuda a analisar e interpretar grandes quantidades de dados.

O que é Ciência de Dados?

As pessoas que trabalham com Ciência de Dados são chamados **Cientistas de Dados**, porém muitas outras carreiras estão associadas a Data Science, como veremos mais adiante.

O que é Ciência de Dados?

O termo Ciência de dados surgiu na EMC (www.emc.com) e vem ganhando popularidade a cada dia.

Ciência de Dados é o processo para extrair informações valiosas a partir de dados. Como estamos vivendo na era do Big Data, a Ciência de dados está se tornando um campo muito promissor para explorar e processar grandes volumes de dados gerados a partir de várias fontes e em diferentes velocidades.

O que é Ciência de Dados?

Ciência de dados é uma grande disciplina em si e consiste em conjuntos de habilidades especializadas, tais como: estatística, matemática, programação, computação e conhecimento de negócios, além de técnicas e teorias, como a análise preditiva, modelagem, engenharia e mineração de dados e visualização.

O que é Ciência de Dados?

A disciplina de Ciência de Dados não evoluiu da noite para o dia

Data Science Academy

O que é Ciência de Dados?

Na verdade, a Ciência de Dados tem estado entre nós há bastante tempo, sob a forma de análise de negócios ou inteligência competitiva, mas somente agora o seu verdadeiro potencial foi percebido e isso se deve em parte ao Big Data.

O que é Ciência de Dados?

O principal objetivo da Ciência de Dados é extrair e interpretar os dados de forma eficaz e apresentá-los em uma linguagem simples e não técnica para os usuários finais e tomadores de decisão.

O que é Ciência de Dados?

Assim, a Ciência de Dados é tudo aquilo relacionado sobre a construção de informações úteis e a capacidade de converter estas informações em produtos ”data-driven”!

Áreas de Conhecimento

Áreas de Conhecimento da Ciência de Dados

Áreas de Conhecimento

Áreas de Conhecimento

Na sua essência, a Ciência de Dados envolve o uso de métodos automatizados (ciência da computação) para analisar enormes quantidades de dados (estatística) e para extrair conhecimento (áreas de negócio) a partir deles.

Áreas de Conhecimento

Áreas de Conhecimento

Áreas de Conhecimento

Perceba que a palavra **chave** em Data Science, não é Data e sim Science.

O que é Ciência de Dados?

Ciência de Dados é a exploração e análise de todos os dados disponíveis, sejam eles estruturados ou não, com o objetivo de desenvolver compreensão, extrair conhecimento e formular ações que gerem resultados.

O que é Ciência de Dados?

A Ciência de Dados está ajudando a criar novos ramos da ciência e influenciando áreas das ciências sociais e humanas.

O que é Ciência de Dados?

E esta tendência deve acelerar nos próximos anos com dados de sensores móveis, instrumentos sofisticados, o crescimento da internet e muito mais.

O que é Ciência de Dados?

Cada vez mais, veremos as universidades oferecendo disciplinas relacionadas a análise de dados, desde de cursos de Engenharia e Ciência da Computação, até Direito e Pedagogia.

O que é Ciência de Dados?

A análise de dados estará presente em todas as áreas de conhecimento.

Dados para Tomada de Decisão

O que é Ciência de Dados?

O que é Ciência de Dados?

Exemplo:

Problema identificado: Clientes cancelam seus planos de telefonia com frequência de 2 a 3 meses após assinatura do plano.

1- Dados e Análise - Dados coletados e analisados, levaram à conclusão que existe uma tendência entre os clientes de operadoras de telefonia em trocar de operadora após o terceiro contato com atendimento ao cliente.

2- Decisão - Os executivos da empresa, baseados em dados, decidiram rever todos os processos de atendimento ao cliente.

3- Ação - a empresa contratou uma consultoria para fornecer treinamento especializado de atendimento ao cliente, passou a monitorar as ligações, criou programas de incentivo aos melhores operadores, etc...

O que é Ciência de Dados?

Dados → Decisões → Ações

O principal objetivo é extrair informações dos dados e transformá-las em conhecimento que possa ser usado para a tomada de decisões

O que é Ciência de Dados?

O que é Ciência de Dados?

O que é Ciência de Dados?

O que é Data Science afinal

O que é Data Science afinal

- ✓ Identificar o problema da área de negócio
- ✓ Compreender o problema (entidades e atributos)
- ✓ Coletar conjuntos de dados (datasets), que representem a entidade
- ✓ Limpar e transformar os dados
- ✓ Compreender os relacionamentos entre os dados
- ✓ Criar modelos que representem os relacionamentos
- ✓ Utilizar os modelos para fazer predições
- ✓ Entregar valor e resultado

O que é Ciência de Dados?

Simples?

Data Science Academy

O que é Ciência de Dados?

Não é necessariamente simples!

data
science
engineering
committee
social
across
center
areas
metrics
funding
scientific
physical
statistics
career
university
one
budget
environment
methodology
engineering
described
studio
faculty
projects
reproducibility
tools
students
open
alternative
see
community
effort
space
working
also
researchers
collaborations
support
scientists
section
project
group
administrative
education
domain
incubator
bridge
escience
director
evaluation
staff
people
participants
collaboration
establish
established
graduate
research
activity
campus
sciences
fellows
activities
professor
impact
year
time
including
environments
long-term
training
participate
paths
software
number
development
program
institute
fellow
courses
postdocs
flagship
ethnography
steering
appendix
theme
computer
programs
new
work
funded

O que é Ciência de Dados?

Normalmente, o processo compreende várias etapas, começando por uma questão de negócio a ser resolvida. Uma vez que você sabe o que você quer analisar, você precisa obter os dados corretos, limpá-los, explorá-los, criar e avaliar um modelo, repetir este ciclo algumas vezes e, finalmente, você está pronto para começar a procurar uma maneira de como comunicar adequadamente seus resultados.

Busque conhecimentos para se tornar um profissional de Analytics.

Isso fará toda a diferença em sua carreira!

Data Science Academy

Ciência De Dados e Big Data

O que são dados?

44,000,000
MESSAGES PROCESSED
486,000
PHOTOS
26
NEW REVIEWS
POSTED ON YELP
120
NEW ACCOUNTS
OPENED ON
LINKEDIN
MORE THAN
140
SUBMISSIONS
ON REDDIT
MORE THAN
2,315,000
SEARCHES
3,125,000
243,055
MORE THAN
21,000,000
MESSAGES SENT
70,000
VIDEO MESSAGES
SHARED
MORE THAN
195,000
MINUTES OF AUDIO CHATTING
ON WECHAT
Share
MORE THAN
3,000,000
ITEMS ARE
SHARED
MORE THAN
18,000
MATCHES MADE
972,000
DAILY SWIPES
ON TINDER
MORE THAN
69,500
HOURS OF
VIDEO WATCHED
ON NETFLIX
NETFLIX
IN
60
SECONDS
GO-Globe
CUSTOM WEB DEVELOPMENT
MORE THAN
150,000,000
E-MAILS ARE SENT
MORE THAN
48,000
APPS DOWNLOADED
ON IPHONE
MORE THAN
95,000
APPS DOWNLOADED
ON ANDROID
YouTube
MORE THAN
430,000
TWEETS SENT
MORE THAN
2,700,000
VIDEO VIEWS AND
139,000 HOURS
OF VIDEO WATCHED
MORE THAN
39,300
HOURS OF MUSIC
LISTENED
AROUND
56,000
PHOTOS
UPLOADED
9,800
ARTICLES PINNED
ON PINTEREST
MORE THAN
280,000
SNAPS SENT
ON SNAPCHAT
MORE THAN
100
NEW DOMAINS
REGISTERED
14 NEW
SONGS ADDED
ON SPOTIFY
MORE THAN
300 HOURS
OF VIDEO ARE UPLOADED
Data Science Academy

Ainda acha pouco

Data Growth
Driven by Unstructured Data
125 Exabytes
Growing 42.5% Per Year
37.4 Exabytes
Growing 22.4% Per Year
79.2%
Unstructured Data
20.8%
Structured Data
Application
Data base
Storage Options
Digital Archive
Backup
2013
2017
ORACLE
NABSHOW
etc

Ciência de Dados e Big Data

2012 → **2.8 Zettabytes**

2020 → **140 Zettabytes**

O que os dados representam?

Dados x Informação

Ciência de Dados e Big Data

Dados são coleções de fatos, tais como números, palavras, medições, observações ou mesmo apenas descrições de coisas.

Ciência de Dados e Big Data

A maioria das pessoas acredita que os termos "dados" e "informação" são intercambiáveis e significam a mesma coisa.

No entanto, há uma diferença distinta entre as duas palavras.

Ciência de Dados e Big Data

Os dados podem ser qualquer caracter, texto, palavras, números, imagens, som ou vídeo e, se não colocados em contexto, significam pouco ou nada para um ser humano.

No entanto, a informação é útil e formatada de uma maneira geral, permitindo que seja entendida por um ser humano.

e Dados e

"DATA IS THE NEW OIL."
Coined in 2006 by Clive Humby, a British data commercialization entrepreneur, this now famous phrase was embraced by the World Economic Forum in a 2011 report, which considered data to be an economic asset, like oil.
From the beginning of recorded time until 2003, we created
5 exabytes of data.
There are nearly as many bits of information in the digital universe as there are stars in our actual universe.
There are 133 million BLOGS on the web.
As of August 2012, there were just over 4 million articles in the English Wikipedia.
English is the dominant language of the web. But by 2014 it will be Chinese, if its current rate of increase continues.
Top languages used on the web (May 2011)
Korean
Russian
French
80% of all humans own a mobile phone of some sort. Out of 5 billion mobiles, 1 billion are smartphones. In Singapore, 84% of citizens are smartphone users.
247 billion EMAILS
Just as a study of activity on Twitter gave residents, family members, and journalists advance warning of details about the devastating earthquake and tsunami in Japan,
high-frequency traders, with the help of computer algorithms, use Big Data to follow trends and to act quickly on their findings.
These specialized algorithms make split-second decisions to buy or sell a commodity.
New cable being laid under the Atlantic will shave 5 milliseconds from the current 65 milliseconds it takes for trading information to travel between New York City and London.
With new fiber-optic cable, the round-trip time between New York and London will be 59.6 milliseconds.
This 5-millisecond saving is worth many millions of dollars to the trading firms who use the cable (and who will pay millions to do so).
How they save 5 milliseconds
The depth of the Atlantic Ocean varies.
The new cable will be on areas of the ocean floor that are up to 1,000 feet shallower than the current fastest cable. By taking a different route, the new cable is shorter, meaning that the time it takes for messages to travel
60% of all humans (5.4 billion people) are active texters. In 2010, 193,000 text messages were sent every second.
10% of all photos ever taken were taken in 2011.

Soluções de Analytics sem dados, são como carros esportivos sem combustível!

Dados estão por todos os lados

TU/e Technische Universiteit Eindhoven University of Technology
Data Science Academy

O que é um dataset?

Conjunto de Dados

O que é um dataset?

- Coleção de observações
- Cada observação é tipicamente chamada de registro
- Cada registro tem um conjunto de atributos que apontam características, comportamentos ou resultados

Datasets

Exemplo: bancos de dados

Exemplo: e-mail

Exemplo: twitter

Cientistas de Dados coletam e utilizam datasets para aprender sobre entidades e prever seu comportamento futuro e possíveis resultados!

Ciência de Dados e Big Data

Formato dos Dados

Fonte dos Dados	Estruturado	Não-estruturado
Interno	• Resultados de pesquisas • Registros de vendas • Medidas de controle de processos • Bancos de dados de sistemas internos (ERP, CRM)	• E-mails, cartas, mensagens de texto • Legendas de vídeos • Comentários de clientes • Mensagens de voz • Imagens / ilustrações • Avaliação de funcionários
Externo	• Likes do Facebook, retweets • Horário de publicação de posts, tweets, updates • Pontuação em sites de classificação	• Conteúdo publicado em redes sociais • Comentários em foruns online • Imagens • Vídeos de câmeras de segurança

E o Big Data?

Ciência de Dados e Big Data

Big Data

Big Data é mais do que uma palavra da moda no mundo dos negócios. Todo mundo deixa “rastros” quando cria uma conta no Facebook, abre conta em banco, faz compras em supermercados, etc. São dados como frequência de compra, consumo médio, número de curtidas em posts de determinado assunto, tweets. Multiplique isso bilhões de vezes e você começa a perceber o que é o Big Data.

E o Big Data?

Big Data é uma coleção de conjuntos de dados, grandes e complexos, que não podem ser processados por bancos de dados ou aplicações de processamentos tradicionais

E o Big Data?

- Empresas de diversos portes estão descobrindo novas maneiras de coletar e usar mais dados e os esforços para tornar o Big Data mais popular serão cada vez mais intensos.

- Agora que as organizações já têm mais conhecimento do que é o Big Data, as questões chave mudaram para quais são as estratégias e as habilidades necessárias e como é possível medir o ROI (Retorno Sobre Investimento).

Ciência de Dados e Big Data

Data
Science

Dados = Big D

Ciência de Dados e Big Data

A Oportunidade do Big Data

ERP CRM BI SCM RH

Mídia Social
Análise de Textos
API's
Reconhecimento de Voz
Processamento de Linguagem Natural
Internet das Coisas

Automação
Saas
PaaS

Dados Internos
Estruturados, previsíveis, permanentes, fáceis de obter

Dados Externos
Não-estruturados, randômicos, voláteis, difíceis de obter

e Dados e

Ciência de Dados e Big Data

E por que é importante aprender a analisar o Big Data, utilizando a Ciência de Dados?

Ciência de Dados e Big Data

Até 2018, haverá um deficit de 140 a 190 mil profissionais com habilidades em análise de dados e mais de 1,5 milhão de gerentes e analistas que saibam usar Big Data de forma efetiva para tomada de decisões.

- McKinsey Global Institute "Big Data Report 2015"

Nós tendemos a superestimar o efeito de uma nova tecnologia no curto prazo e subestimar o efeito no longo prazo.

- Roy Amara

Ciência de Dados e Big Data

As profissões relacionadas a análise de dados, estão entre as mais promissoras na próxima década.

Em Deus nós acreditamos. Todos os outros devem trazer dados.

Willian E. Deming

Ciência De Dados e Estatística

Estatística e Data Science são a mesma coisa?

NÃO

Matemática e Estatística
Data Science
Ciência da Computação
Conhecimento das áreas de negócio

- Microsoft Power BI Para Data Science (Gratuito)
- Formação Cientista de Dados
- Formação Inteligência Artificial

O que é Estatística?

A Estatística fornece técnicas e métodos de análise de dados que auxiliam o processo de tomada de decisão nos mais variados problemas onde existe incerteza.

- Bioestatística
- Contabilometria
- Controle de qualidade
- Estatística comercial
- Estatística econômica
- Estatística de engenharia
- Estatística física
- Estatística populacional
- Estatística psicológica
- Estatística social (para todas as ciências sociais)
- Geoestatística
- Pesquisa operacional
- Análise de processo e quimiometria

Data Science Academy

As três grandes áreas da Estatística

- Probabilidade
- Estatística Descritiva
- Estatística Inferencial

Probabilidade

Estudo da aleatoriedade e da incerteza

Estatística Descritiva

Utiliza métodos para coleta, organização, apresentação, análise e síntese de dados obtidos em uma população ou amostra.

Estatística Inferencial

É o processo de estimar informações sobre uma população a partir dos resultados observados em uma amostra.

É verdade que algumas pessoas pensam que as estatísticas mentem.

Ou, como dizem alguns:

“Os números dizem qualquer coisa quando bem torturados”

Mas a verdade é que qualquer ciência produz resultado contrário ao desejado, quando é mal aplicada.

Portanto, as estatísticas “mentem” apenas quando estão erradas ou, no mínimo, estão sendo mal interpretadas.

Todo dado se refere a uma variável.

A Estatística trata dados.

Portanto, a Estatística trabalha com variáveis, que podem assumir diferentes valores, em diferentes unidades.

Data Science Academy

O que é Estatística?

Estatística é a **ciência**, parte da **Matemática Aplicada**, que fornece métodos para **coletar**, **descrever**, **analisar**, **apresentar** e **interpretar** dados, para a utilização dos mesmos na **tomada de decisões**.

Big Data Analytics é o termo que se refere a análise estatística de grandes quantidades de dados, para que se possa extrair informação relevante para a compreensão da situação atual e a tomada de decisões.

Formação Cientista de Dados

- **Big Data Analytics com R e Microsoft Azure Machine Learning**
- **Big Data Real-Time Analytics com Python e Spark**
- Engenharia de Dados com Hadoop e Spark
- **Machine Learning**
- **Business Analytics**
- Visualização de Dados e Design de Dashboards
- Preparação Para a Carreira de Cientista de Dados

Qual a Relação Entre Ciência de Dados e Estatística?

As teorias estatísticas servem como base para tudo, desde a criação de perfis de passageiros em uma era de ameaças terroristas até à eficácia de novos programas para reduzir a taxa de erros hospitalares.

A Estatística pode ajudar a avaliar se o sucesso de um fundo de investimentos é genuíno ou devido ao acaso, pode ajudar a prever se um determinado assinante vai cancelar sua assinatura este ano ou se uma reivindicação de seguro é fraudulenta.

Vivemos uma era em que a ciência deve prevalecer sobre o empirismo, em que a lógica deve prevalecer sobre o ”achismo”.

A Ciência de Dados utiliza os métodos estatísticos para explorar e analisar dados, fazer inferências e buscar padrões em meio a incertezas, tudo isso em novas abordagens e com o auxílio da Ciência da Computação.

Ciência de Dados x Análises Estatísticas

A razão para a necessidade crescente desta nova abordagem está relacionada ao Big Data, que demanda o uso de diferentes tecnologias à análise estatística.

Os Cientistas de Dados possuem 3 características principais:

1. Eles têm um forte conhecimento de Estatística e aprendizado de máquina, pelo menos o suficiente para evitar má interpretação de correlação e causalidade.

2. Eles têm habilidades de informática para usar uma linguagem de programação (como R ou Python) para facilitar o trabalho de análise.

3. Eles podem visualizar e resumir seus dados e sua análise de uma maneira que seja significativa para alguém menos familiarizado com os dados, baseado em sua experiência de áreas de negócio.

A Estatística é parte fundamental do trabalho do Cientista de Dados

A American Statistical Association (ASA) divulgou recentemente uma declaração sobre o papel da Estatística na Ciência de Dados. O Presidente da ASA, David Morganstein, deu esta declaração no seu comunicado de imprensa:

"Através desta declaração, a ASA e seus membros reconhecem que a ciência de dados abrange mais do que estatística, mas ao mesmo tempo também reconhece que a estatística desempenha um papel fundamental no rápido crescimento deste campo. É nossa esperança que esta declaração possa reforçar a relação da estatística com a ciência de dados e ainda fomentar relacionamentos mútuos de colaboração entre todos os contribuintes na ciência de dados."

Fonte: http://www.amstat.org/newsroom

A declaração evidencia que Estatística é fundamental para a Ciência de Dados, juntamente com gestão de banco de dados, sistemas distribuídos e paralelos, computação, matemática e programação. A sua utilização neste campo emergente, capacita pesquisadores para extrair conhecimento e obter melhores resultados de grandes projetos. A declaração também incentiva a colaboração máxima e multifacetada entre estatísticos e cientistas de dados para maximizar o potencial da ciência de dados.

Em resumo, a Estatística desempenha um papel fundamental dentro da Ciência de Dados. Porém, a Ciência de Dados compreende outras áreas de conhecimento, como já vimos anteriormente.

Ciência de Dados também é uma arte!

Ciência de Dados também é uma arte!

A Ciência de Dados não é apenas uma ciência ou uma técnica, é também uma "**arte**". Ciência de dados é a arte de ouvir as suas intuições, enquanto enfrenta enorme quantidade de dados, classificando, avaliando e buscando as conclusões.

Ciência de Dados também é uma arte!

Nem todo mundo é abençoado com esta arte!

Ciência de Dados também é uma arte!

Cientistas de Dados precisam ser realmente criativos em visualizar os dados de várias formas gráficas e apresentar os dados altamente complexos de uma forma muito simples e amigável!

Ciência de Dados também é uma arte!

Se um Cientista de Dados é capaz de converter aterrorizantes Petabytes de dados estruturados, bem como dados não-estruturados (imagens, vídeos, arquivos de log, etc...) em um formato mais fácil e simples, ele é um - "**artista**"!

Ciência de Dados também é uma arte!

Afinal, apenas um Cientista de Dados mais hábil pode gerenciar o banco de dados do McDonalds, gerar inisghts dos bilhões de vídeos carregados no Youtube, gerenciar grandes volumes de dados da GE ou ainda gerenciar os dados relativos às milhares de amostras de sangue de pacientes ou dados não estruturados gerados a partir de raios-X!

O Aprendizado de Estatística

Para muitas pessoas, o aprendizado de Estatística pode ser meio obscuro. Isso se deve aos métodos de ensino, muitas vezes artificiais e focados apenas em fórmulas matemáticas, sem qualquer ligação com o mercado de trabalho ou com nossas vidas diárias.

O Aprendizado de Estatística

Entretanto, o rápido crescimento da Ciência de Dados e da análise de dados em geral, colocou a Estatística no centro das atenções, trazendo novas formas de ensinar e aprender Estatística.

Aqui na DSA acreditamos que a melhor forma de aprender é fazendo! Por isso oferecemos cursos que aliam teoria e prática na medida certa, com conteúdo atual e voltado ao mercado de trabalho! E tudo 100% em português!

Não espere até que você entenda completamente o que você está fazendo, antes de começar a aplicar seu conhecimento.

A compreensão vem à medida que você aplica o conhecimento.

Quando aprendemos a falar, nós não esperamos até compreender a estrutura gramatical do idioma, antes de dizer qualquer coisa.

Quanto mais você aplica o conhecimento, mais você aprende!

Em uma análise estatística, o contexto é vital e muito interessante.

Você precisa entender o problema que deu origem à investigação e coleta de dados (por isso o conhecimento de áreas de negócios é tão importante). O contexto é o que faz cada investigação estatística diferente.

Os Cientistas de Dados muitas vezes trabalham ao lado de outros pesquisadores em áreas como Medicina, Psicologia, Biologia, Geologia, Marketing, Finanças, Contabilidade, E-commerce, etc..,que fornecem o plano de fundo contextual para o problema.

É importante aprender Estatística, para se tornar um Cientista de Dados?

*** SIM ***

E por isso a **Data Science Academy** oferece a Formação Cientista de Dados, um programa com mais de 1.200 aulas em vídeo, mais de 320 scripts comentados linha a linha, mais de 200 manuais em pdf, 26 projetos, suporte em até 24 horas, forum exclusivo, networking de qualidade com alunos de alto nível e todo conteúdo 100% em português!

E você ainda pode estudar do seu smartphone ou tablet com nossas apps gratuitas para iOS e Android.

Data Science Academy

O futuro é aqui.

Introdução à Ciência de Dados 2.0

Aprendizado de Máquina (Machine Learning)

O Que é Aprendizado de Máquina?

O que é Aprendizado?

Aprendizado é a capacidade de se adaptar, modificar e melhorar seu comportamento e suas respostas, sendo portanto uma das propriedades mais importantes dos seres ditos inteligentes, sejam eles humanos ou não.

Estamos tentando reproduzir o processo de aprendizado de seres humanos em máquinas, através de algoritmos de Machine Learning.

Machine Learning é um subcampo da Inteligência Artificial que permite dar aos computadores a habilidade de aprender sem que sejam explicitamente programados para isso.

Machine Learning ou Aprendizado de Máquina é um método de análise de dados que automatiza o desenvolvimento de modelos analíticos. Usando algoritmos que aprendem a partir de dados, o aprendizado de máquina permite que os computadores encontrem insights ocultos sem que sejam explicitamente programados para procurar algo específico.

Since an early flush of optimism in the 1950s, smaller subsets of artificial intelligence – first machine learning, then deep learning, a subset of machine learning – have created ever larger disruptions.

BIG DATA
LARGE
RESEARCH
SCIENCE
INTERNET
INFORMATION
BUSINESS
SEARCH
TIMES
PETABYTES
MANAGEMENT
PROCESS
NEW
SIZE
SOFTWARE
ANALYSIS
MILLION
BILLION
GARTNER
WORLD
SENSOR
NETWORKS
ABOUT
EVERY
SETS
FUTURE
YEARS
USE
WORK
DEVELOPMENT
GOVERNMENT
ECONOMIC
DUE

Tipos de Aprendizagem

Tipos de Aprendizagem

Aprendizagem Supervisionada

Aprendizagem Supervisionada

Aprendizagem Supervisionada

Dados | Amostras de Dados | Algoritmo | Modelo Treinado | Teste e Validação | Produção

Os algoritmos de aprendizado supervisionado fazem previsões com base em um conjunto de exemplos.

Aprendizagem Supervisionada

Classificação

Regressão

Aprendizagem Não Supervisionada

Alguns sistemas de recomendação que você encontra na internet sob a forma de automação de marketing são baseados neste tipo de aprendizagem.

O **Aprendizado Não Supervisionado** é usado com dados que não possuem rótulos históricos, ou seja, nós não temos variáveis target (as variáveis de saída) para serem estimadas.

O algoritmo não recebe durante o treinamento, os possíveis resultados e ele deve descobrir por si próprio. O objetivo é explorar os dados e encontrar alguma estrutura neles.

Aplicações da Ciência de Dados

Aplicações da Ciência de Dados

O potencial de aplicação da Ciência de Dados é imenso.

E este é um dos fatores que tem feito com que Data Science tenha atraído tanto o interesse de empresas, profissionais e do mercado em geral.

Data Science

Aplicações da Ciência de Dados

Vejamos alguns exemplos.

- Detecção de fraudes
- Carros automatizados
- Melhores sistemas de monitoramento
- Detecção e prevenção de epidemias
- Detecção de terremotos
- Educação customizada, por demanda e online
- Medicamentos customizados, baseados no histórico de cada paciente
- Processo otimizado de iluminação residencial, industrial e pública
- Sistemas de busca mais eficientes
- Geo-marketing através de smartphones
- Marketing personalizado
- Combate ao crime e ao terrorismo

Aplicações da Ciência de Dados

Temos visto nos últimos anos, investimentos em toda a infraestrutura de negócios, o que tem gerado habilidade de coletar todos os tipos de dados através das empresas.

Aplicações da Ciência de Dados

Atualmente, cada aspecto de um negócio está aberto para a coleta de dados

Operações
Manufatura
Gestão de Logística
Comportamento do Cliente
Processos e Procedimentos
Performance de Campanhas de Marketing

Aplicações da Ciência de Dados

Em paralelo, a informação está vastamente disponível, como tendências de mercado, novas tecnologias e movimento dos competidores.

Aplicações da Ciência de Dados

Esta disponibilidade de dados, tem aumentado o interesse por métodos de extração de informação útil e conhecimento dos dados

E este é o reino da Ciência de Dados

Aplicações da Ciência de Dados

A Ciência de Dados pode ser utilizada em praticamente todas as atividades humanas, desde que dados sejam gerados e possam ser coletados, incluindo aplicações avançadas, tais como:

- Sistemas de Inteligência Artificial
- Agentes Virtuais Inteligentes
- Robôs Advogados
- Robôs Investidores
- Robôs Cirurgiões
- Sistemas de Tomada de Decisão

Exemplos de Aplicação da Ciência de Dados

Aplicações da Ciência de Dados

- Servidores de dados com 4 Petabytes de dados
- Cada venda é registrada
- Aproximadamente 267 milhões de transações por dia, nas 6000 lojas em todo o mundo
- Análise de Dados focada na avaliação da efetividade de estratégias de preço e campanhas de marketing
- Busca de melhoria na sua gestão logística e de inventário

Aplicações da Ciência de Dados

amazon.com®

- Personalização da experiência de compra online
- Cada cliente possui sua própria loja, baseada nas suas preferências
- Influência das avaliações de outros usuários, nas decisões de compra

Aplicações da Ciência de Dados

- A análise de cada uma das transações realizadas pelo banco, nos mais de 100 países em que opera, permite a geração de insights relacionados a investimentos, mudanças de mercado, padrões de operação e condições econômicas

Aplicações da Ciência de Dados

Outros exemplos:

- Netflix
- Mídias Sociais (Facebook, LinkedIn, Twitter)
- Web Apps (Uber, AirBnB)
- Planejamento urbano (Cidades Europeias)
- Astrofísica (Nasa)
- Saúde Pública (Hospitais Americanos)
- Esportes (NFL)

Aplicações da Ciência de Dados

Educação

- Orientação em tempo real para que o estudante não abandone a escola
- Personalização do processo de aprendizagem
- Monitoramento do estudante na sua vida acadêmica

Aplicações da Ciência de Dados

Educação

Imagine se escolas e faculdades conhecessem melhor seus estudantes, suas limitações, forças e fraquezas e com isso pudessem direcionar e instruir cada um deles com formas personalizadas de aprendizado mais desafiador.

Imagine se estes dados pudessem ser cruzados com indicadores sócio-econômicos, renda familiar, previdência social, etc..., gerando com isso informação para governos e instituições.

Aplicações da Ciência de Dados

Varejo

Esta talvez seja uma das áreas mais beneficiadas pela Ciência de Dados, devido a diversidade de dados gerados.

- Análise de sentimento da marca
- Recomendação individualizada de produtos
- Retenção e satisfação do cliente

Aplicações da Ciência de Dados

Telecom

Desenvolver novos produtos já que os dispositivos móveis estão produzindo uma grande quantidade de dados sobre como, por que, onde e quando estão sendo usados. Estes dados são extremamente valiosos, mas devido ao volume e variedade fica difícil ingerir, armazenar e analisar estes dados que podem resultar em **novos grandes *insights***.

- Análise de registro de ligações
- Geolocalização de ligações
- Ofertas personalizadas de produtos
- Racionalização dos custos

Aplicações da Ciência de Dados

Saúde

- Monitoramento de sinais vitais
- Redução de taxa de retorno de pacientes
- Medicamentos personalizados

Aplicações da Ciência de Dados

Financeiro

- Detecção de fraudes
- Análise de risco
- Análise de perfil
- Geolocalização

Aplicações da Ciência de Dados

Governos

- Fornecimento personalizado de serviços públicos
- Mapeamento de segurança pública
- Redução e racionalização de gastos

Aplicações da Ciência de Dados

Percebeu por que a Ciência de Dados cresce em alta velocidade?

Nunca o mundo gerou tantos dados e por isso mesmo, nunca tivemos tantas oportunidades como agora!

Pense nisso!

Business Intelligence não é a mesma coisa que Data Science.

Aplicações da Ciência de Dados

Business Intelligence não é a mesma coisa que Data Science

Business Intelligence e Data Science tem muita coisa em comum e Cientistas de Dados e Analistas de BI são como primos.

Aplicações da Ciência de Dados

Business Intelligence não é a mesma coisa que Data Science

Ambos usam dados com o mesmo objetivo, mas a sua abordagem, tecnologia e função diferem de diversas maneiras.

Aplicações da Ciência de Dados

Business Intelligence

O objetivo do BI é converter dados brutos em insights de negócio, que os líderes empresariais e gestores possam usar para tomar decisões.

Ciência de Dados

A Ciência de Dados também converte dados brutos em insights de negócio, mas aplica metodologia científica para exploração dos dados, testes de hipótese, modelagem estatística e aprendizado de máquina.

Aplicações da Ciência de Dados

Análise

Descritiva	O que aconteceu?
Diagnóstica	Por que isso aconteceu?
Preditiva	O que acontecerá?
Prescritiva	O que deve ser feito?

BI Tradicional

Ciência de Dados

Aplicações da Ciência de Dados

DataOps

Aplicações da Ciência de Dados

DataOps é a operação da infraestrutura necessária para suportar a quantidade, velocidade e variedade de dados disponíveis na empresa, que hoje é radicalmente diferente das abordagens de gerenciamento de dados tradicionais.

Aplicações da Ciência de Dados

A natureza do DataOps abraça a necessidade de gerenciar fontes de dados muito diferentes e gerados em tempo real, permitindo que o Cientista de Dados tenha a sua disposição a infraestrutura necessária para fazer seu trabalho!

Aplicações da Ciência de Dados

Aplicações da Ciência de Dados

DevOps → DataOps

Aplicações da Ciência de Dados

DataOps reconhece a natureza interligada da engenharia de dados, integração de dados, qualidade de dados e segurança de dados e tem como objetivo ajudar uma empresa a entregar rapidamente os dados que poderão acelerar a análise e permitir resultados antes impossíveis.

Aplicações da Ciência de Dados

Em resumo, DataOps, é um conjunto de melhores práticas que visam tornar eficiente a coordenação entre a Ciência de Dados e as operações de dados e tornou-se assim um assunto fundamental para qualquer organização de TI que queira sobreviver e prosperar em um mundo onde inteligência de negócios em tempo real é uma necessidade competitiva.

Aplicações da Ciência de Dados

A Ciência de Dados é uma disciplina extremamente importante atualmente

Aplicações da Ciência de Dados

Mas essa disciplina só é útil na medida em que ela possa ser executada de forma confiável e eficiente

Aplicações da Ciência de Dados

E para que isso aconteça, você precisa do DataOps para que seus DBAs, Cientistas de Dados, Desenvolvedores, Infraestrutura e Operações estejam em harmonia e focados na mesma direção.

Aplicações da Ciência de Dados

Aplicações da Ciência de Dados

Data Lake é um termo recente, criado pelo CTO do Pentaho, James Dixon, para descrever um componente importante no universo da Ciência de Dados e do Big Data.

Aplicações da Ciência de Dados

Data Lake

A ideia é ter um único repositório dentro da empresa, para que todos os dados brutos estejam disponíveis a qualquer pessoa que precise fazer análise sobre eles.

Aplicações da Ciência de Dados

Data Lake

Aplicações da Ciência de Dados

Aplicações da Ciência de Dados

Os Data Lakes armazenam os dados em seu formato bruto, sem qualquer processamento e sem governança.

Aplicações da Ciência de Dados

Aliás, apesar das soluções comerciais, Data Lake é um conceito e não uma tecnologia.

Aplicações da Ciência de Dados

O Data Lake em essência, é uma ***estratégia de armazenamento de dados.***

Aplicações da Ciência de Dados

Quando se ouve falar sobre um ponto único para reunir todos os dados que uma organização deseja analisar, imediatamente se imagina a noção de Data Warehouse e Data Mart.

Aplicações da Ciência de Dados

Mas há uma distinção fundamental entre Data Lake e Data Warehouse.

Aplicações da Ciência de Dados

Data Lake

O Data Lake armazena dados brutos, sob qualquer forma do jeito que foram coletados na fonte de dados.

Aplicações da Ciência de Dados

Data Lake

Não há suposições sobre o esquema dos dados e cada fonte de dados pode usar qualquer esquema. Cabe aqueles que vão analisar os dados, dar sentido a esses dados para o propósito ao qual a análise se destina.

Aplicações da Ciência de Dados

Data Warehouse

Em contrapartida, o Data Warehouse tende a usar a noção de um único esquema para todas as necessidades de análise, o que se torna impraticável em muitas situações relacionadas a Big Data.

Aplicações da Ciência de Dados

Data Warehouse

Os dados são limpos e organizados antes do armazenamento, fazendo com que os dados estejam disponíveis para uso e análise, assim que são armazenados.

Aplicações da Ciência de Dados

Ao mudar o foco para o armazenamento dos dados brutos, isso coloca a responsabilidade sobre os Cientistas de Dados.

Aplicações da Ciência de Dados

O Data Lake passa a ser a fonte para a Ciência de Dados

Aplicações da Ciência de Dados

Enterprise Data Hub

Aplicações da Ciência de Dados

Um Enterprise Data Hub é um modelo de gerenciamento de Big Data que usa uma plataforma Hadoop como o repositório de dados central.

Aplicações da Ciência de Dados

O objetivo de um Enterprise Data Hub é permitir que a empresa tenha uma fonte de dados centralizada e unificada que possa fornecer rapidamente informações a diversos usuários de negócio, apoiando a tomada de decisão.

Aplicações da Ciência de Dados

Todos os aplicativos de analytics se conectam ao EDH para extrair as informações para análise.

Aplicações da Ciência de Dados

A introdução de um Enterprise Data Hub no cerne da arquitetura de informação de uma empresa, promove a centralização de todos os dados, em todos os formatos, disponíveis para **todos** os usuários de negócio, com total fidelidade e segurança e custo até 99% menor por Terabyte, em comparação com um Data Warehouse tradicional.

Aplicações da Ciência de Dados

Um Enterprise Data Hub serve como um repositório flexível para coletar e manter dados de forma ilimitada, seja para fins de conformidade ou para aplicações sofisticadas, como detecção de anomalias em tempo real.

Aplicações da Ciência de Dados

O conceito de **EDH** (Enterprise Data Hub), não chega a ser exatamente novo, pois já existem os **EDWs** (Enterprise Data Warehouses).

Aplicações da Ciência de Dados

O desafio está em mover a empresa de um modelo tradicional de gestão de dados, para um modelo voltado ao Big Data e suas particularidades como volume, velocidade e variedade, permitindo assim a plena utilização da Ciência de Dados.

Aplicações da Ciência de Dados

O Enterprise Data Hub inclui:

- Reservatório de Dados ou "Data Lake
- Exploração do Big Data
- Fácil acesso aos dados
- Armazenamento de dados em formato nativo
- Fonte para projetos de Data Science

Ciclo de Vida de Projetos de Data Science

Ciclo de Vida de Projetos de Data Science

Todo projeto de Data Science deve começar com o objetivo, ou seja, as questões que precisam ser respondidas.

Ciclo de Vida de Projetos de Data Science

Depois de formuladas as questões, busca-se os dados que ajudarão a respondê-las.

Ciclo de Vida de Projetos de Data Science

A maturidade da disciplina de Gestão de Projetos, permite que boas práticas de gerenciamento de projetos, como as definidas pelo PMI (Project Management Institute), possam ser utilizadas em projetos de Data Science.

Ciclo de Vida de Projetos de Data Science

Não há porque reinventar a roda. Conhecimento e experiência adquiridos em Gestão de Projetos, serão fundamentais para o sucesso de projetos de Data Science.

Ciclo de Vida de Projetos de Data Science

Ciclo de Vida de Projetos de Data Science

Inicialmente, é vital compreender o problema a ser resolvido.	Isso pode parecer óbvio, mas projetos de negócios raramente vem empacotados de forma clara.
Avaliar e reavaliar o problema e desenhar uma solução, é um processo iterativo de descoberta.	Na prática, o maior desafio em Ciência de Dados, está na identificação do problema.

Ciclo de Vida de Projetos de Data Science

Identificar o problema é uma das tarefas principais em um projeto de Data Science

Atenção!!

Não subestime esta etapa.

A identificação correta do problema é a chave para o sucesso de projetos de Ciência de Dados.

Ciclo de Vida de Projetos de Data Science

Identificar o problema é uma das tarefas principais em um projeto de Data Science

E diversas técnicas analíticas podem ser usadas no processo:

- Estatística
- Busca em bancos de dados (Database Query)
- Data Warehousing
- Análise de Regressão
- Machine Learning e Data Mining

Ciclo de Vida de Projetos de Data Science

Vamos fazer um simples exercício:

Vamos elaborar algumas perguntas para identificar o problema e a possível solução a ser adotada

Ciclo de Vida de Projetos de Data Science

Quais são os clientes mais rentáveis?

Esta pergunta poderia facilmente ser respondida com uma consulta simples a um banco de dados, usando linguagem SQL.

Ciclo de Vida de Projetos de Data Science

Existe diferença entre os clientes mais rentáveis e a média dos clientes?

Aqui poderíamos usar a Estatística, realizando um teste de hipótese para confirmar ou não nossa tese que de os clientes mais rentáveis possuem diferença em relação à média dos clientes.

Ciclo de Vida de Projetos de Data Science

Algum cliente em particular estará no grupo dos mais rentáveis? Quanto de faturamento posso esperar vindo deste cliente?

Aqui poderíamos utilizar Data Mining, para examinar o histórico dos clientes e criar modelos preditivos de geração de lucro por cliente. Estes modelos poderiam ser aplicados automaticamente usando Machine Learning.

Ciclo de Vida de Projetos de Data Science

A identificação do problema a ser resolvido, vai impactar todo o projeto de Data Science, não só com relação ao custo, mas também com relação aos recursos que serão adotados.

Ciclo de Vida de Projetos de Data Science

E quanto mais complexas as perguntas, mais a Ciência de Dados pode ajudar nas respostas!!

Ciclo de Vida de Projetos de Data Science

Não existe a melhor tecnologia

Existe aquela que melhor atende o requisito do cliente

Ciclo de Vida

Ciclo de Vida de Projetos de Data Science

Ciclo de Vida de Projetos de Data Science

Preparação	Engenharia de Dados	Analytics	Produção
Definição do objetivo	Aquisição	Análise Exploratória de Dados	Construção de produtos de dados
Compreensão do problema	Limpeza	Inferências	Operacionalização de alimentação de dados
Conhecimento dos dados	Transformação	Modelagem	Melhoria contínua
	Enriquecimento	Predição	
	Persistência	Comunicação	

Ciclo de Vida de Projetos de Data Science

Ciclo de Vida de Projetos de Data Science

A qualidade dos seus outputs será determinada pelas qualidade dos seus inputs

Bons resultados dependem da formulação de boas questões

Ciclo de Vida de Projetos de Data Science

Produtos gerados em projetos de Data Science

- Relatórios
- Narrativas
- Apresentações
- Web Sites
- Aplicativos

- Facilidade de Uso
- Reprodutibilidade
- Documentação
- Conclusões Concisas

Ciclo de Vida de Projetos de Data Science

Cultura Data-Driven

Ciclo de Vida de Projetos de Data Science

Cultura Data-Driven

Uma empresa com cultura orientada a dados, reconhece a importância do Big Data e como a análise destes dados, pode levar a empresa a criar produtos e serviços personalizados e totalmente aderentes as necessidades do público.

Ciclo de Vida de Projetos de Data Science

A mudança de paradigma será inevitável. Aliás, já estamos vivenciando esta mudança. Terá sucesso, quem melhor se adaptar a ela.

Ciclo de Vida de Projetos de Data Science

Os dados já fazem parte do negócio.

Ciclo de Vida de Projetos de Data Science

Exemplos de Projetos de Data Science

- Análise de Texto das avaliações de itens comprados online
- Previsão de comportamento de usuário
- Previsão de desempenho de esportistas ou clubes esportivos
- Análise de Marketing e Mídia Social para redes de hotéis
- Previsão de Crimes e locais de ocorrências
- Análise e previsão de taxas de inadimplência
- Sistemas de recomendação
- Previsão de atrasos em voos
- Análise e Classificação de arquivos potencialmente maliciosos

Ciclo de Vida de Projetos de Data Science

Projetos de Data Science são cada vez mais comuns e estão mudando a forma como as empresas fazem negócio.

Estão mudando a vida das pessoas.

Carreiras em Data Science

Carreiras em Data Science

Com a explosão da análise de dados e do Big Data, a busca por profissionais capazes de extrair, analisar e gerar insights dos dados, não para de crescer.

Carreiras em Data Science

Veja este gráfico do site de empregos Indeed

Carreiras em Data Science

Gráfico do Google Trends

Carreiras em Data Science

Crescimento do volume de dados x Crescimento da formação de Analistas de Dados

O armazenamento de dados aumenta 28% ao ano, enquanto a formação de profissionais de Análise de Dados aumenta a 5,7% ao ano.

Carreiras em Data Science

Crescimento do volume de dados x Crescimento da formação de Analistas de Dados

Neste ritmo, o que vai acontecer daqui 5 anos?

Carreiras em Data Science

Crescimento do volume de dados x Crescimento da formação de Analistas de Dados

O mercado de trabalho não terá profissionais suficientes para suprir a demanda!!!!

Carreiras em Data Science

Até 2024, estima-se que 35 Zettabytes (1 ZB = 10^{21} bytes) de dados estejam circulando na rede.

Carreiras em Data Science

Ou seja, as empresas carecem cada vez mais da experiência daqueles que trabalham com tecnologias como o Big Data - leia-se profissionais experts em ciência da computação, estatística, matemática e domínio de negócios.

Carreiras em Data Science

A partir daí, extrair o máximo de informações relevantes e decifrá-las com velocidade, torna-se a chave para transformar dados em novos negócios.

Carreiras em Data Science

A explosão na busca por profissionais de dados tem levado a criação de novas carreiras, bem como a reinvenção de outras.

Carreiras em Data Science

Por mais que os conceitos de análise de dados existam a décadas, as tecnologias que surgiram recentemente, permitem fazer coisas que não eram possíveis antes, sem falar no fato que o volume de dados gerado pela humanidade nunca foi tão grande.

Carreiras em Data Science

Na prática, todas as profissões envolvem a coleta e análise de dados.

Carreiras em Data Science

Carreiras em Data Science

Vejamos as carreiras que mais estão sendo requisitadas, principalmente pelas empresas que já começaram seus projetos de Big Data e que precisam de profissionais para a Ciência de Dados.

Carreiras em Data Science

De forma geral, podemos identificar três perfis básicos de profissionais engajados em Data Science e Big Data:

Carreiras em Data Science

1) Cientistas de dados

Profissionais capacitados em estatística, ciência da computação e/ou matemática capazes de analisar grandes volumes de dados e extrair deles insights que criem novas oportunidades de negócio.

Carreiras em Data Science

2) Analistas de negócio

Conhecendo bem o negócio em que atuam conseguem formular as perguntas corretas, analisar as respostas e tomar decisões estratégicas e táticas que alavanquem novos negócios ou aumentem a lucratividade da empresa. Esta função tende a ser acoplada às funções do Cientista de Dados.

Carreiras em Data Science

3) Profissionais de tecnologia

Cuidam da infraestrutura e do suporte técnico para suportar Big Data. O aparato tecnológico de Big Data não é muito comum em empresas tipicamente comerciais, pois demanda expertise em gerenciar hardware em clusters de alta performance.

Carreiras em Data Science

Entretanto, nos próximos anos viveremos uma escassez destes profissionais, não só no Brasil, mas no mundo todo.

Carreiras em Data Science

Esta escassez ao mesmo tempo em que abre muitas perspectivas profissionais para os que abraçarem a função, também atuará como um entrave, pois dificultará às empresas usarem Big Data com eficiência.

Carreiras em Data Science

Recentes pesquisas estimam que por volta de 2018, Big Data demandará cerca de 4,4 milhões de profissionais em todo o mundo e que apenas 1/3 destes cargos poderá ser preenchido com as capacitações disponíveis hoje em dia.

Carreiras em Data Science

Uma pesquisa mundial da IBM corrobora estes dados, mostrando que apenas uma em dez organizações acredita que tenha profissionais com as capacitações necessárias e que três em cada quatro estudantes e professores reportam que existe um gap de moderado a grande entre o que é ensinado hoje e o que o mercado de trabalho realmente necessita.

Carreiras em Data Science

Carreiras em Data Science

Analista de Negócios

Carreiras em Data Science

Analista de Negócios

- Estabelecer os objetivos e o âmbito de sistemas de negócios e de TI
- Identificar problemas organizacionais e conceber soluções orientadas a dados
- Realizar análises estatísticas, pesquisas, oficinas de formação e testes
- Recomendar mudanças nos processos, pessoal ou ofertas de produtos para tornar os departamentos internos mais eficientes
- Conceber novos sistemas ou alterar os existentes
- Fazer recomendações específicas de TI e apoiar a sua implementação
- Agir como um elo de ligação entre os gestores e equipes técnicas
- Propor suas decisões baseadas em dados

Curso recomendado na DSA: **Business Analytics**

Análise de Dados

Carreiras em Data Science

Analista de Dados

- Trabalhar com as equipes de TI, gestão e / ou Cientistas de Dados para determinar os objetivos organizacionais
- Coletar dados de fontes primárias e secundárias
- Realizar limpeza nos dados e descartar informações irrelevantes
- Analisar e interpretar os resultados utilizando ferramentas estatísticas e técnicas convencionais
- Identificar tendências, correlações e padrões em conjuntos de dados complexos
- Identificar novas oportunidades para melhoria de processos
- Concepção, criação e manutenção de bancos de dados relacionais e NoSQL e sistemas de dados
- Resolver problemas de código e questões relacionadas a dados
- Dominar linguagens (R, Python, SQL) e softwares de análise de dados (SAS, Tableau, Qlik)

Cursos recomendados na DSA: **Big Data Analytics com R e Azure e Big Data Real-Time Analytics com Python e Spark**

Carreiras em Data Science

Arquiteto de Dados

- Colaborar com as equipes de TI e gestão para elaborar uma estratégia de dados que atenda os requisitos da empresa
- Criar um inventário de dados necessários para implementar a arquitetura
- Pesquisar novas oportunidades de aquisição de dados
- Identificar e avaliar as atuais tecnologias de gerenciamento de dados
- Criar um fluxo de dados dentro da empresa
- Desenvolver modelos de dados
- Projetar, documentar, construir e implantar arquiteturas e aplicações de banco de dados (por exemplo, grandes bancos de dados relacionais e NoSQL)
- Integrar a funcionalidade técnica (por exemplo, escalabilidade, segurança, desempenho, recuperação de dados, confiabilidade, etc.)
- Implementar medidas para assegurar a precisão dos dados e acessibilidade

Cursos recomendados na DSA: **Arquiteto de Soluções AWS, Engenharia de Dados com Hadoop e Spark ou Formação Engenheiro de Dados**

Engenheiro de Dados

Carreiras em Data Science

Engenheiro de Dados

- Projetar, construir, instalar, testar e manter sistemas de gerenciamento de dados altamente escaláveis
- Construir algoritmos de alto desempenho, protótipos, modelos preditivos e provas de conceito
- Pesquisar a aquisição de dados e novos usos para os dados existentes
- Desenvolver processos de conjuntos de dados para modelagem de dados, mineração e produção
- Integrar novas tecnologias de gerenciamento de dados e ferramentas de engenharia de software nas estruturas existentes
- Criar componentes personalizados de software e aplicações analíticas
- Empregar uma variedade de linguagens e ferramentas
- Instalar e atualizar os procedimentos de recuperação de desastres
- Recomendar formas de melhorar a confiabilidade dos dados, eficiência e qualidade
- Dominar tecnologias como Hadoop, Spark e Cassandra

Cursos recomendados na DSA: **Formação Engenheiro de Dados**

Administrador de Banco de Dados

Carreiras em Data Science

Administrador de Banco de Dados

- Suporte técnico aos bancos de dados existentes
- Personalização de bancos de dados comerciais para necessidades específicas
- Planejamento e projeto de bancos de dados para necessidades específicas
- Solução de problemas para atender às necessidades dos clientes
- Desenvolvimento de Banco de Dados para uma ampla variedade de aplicações
- Supervisão da instalação de novos SGBD
- Criar procedimentos de backup, restauração e recuperação de desastres
- Atuar com bancos de dados relacionais e não relacionais

Cursos recomendados na DSA: **Design e Implementação de Data Warehouses, Gerenciamento de Dados com MongoDB e SQL Para Big Data**

Estatístico

Carreiras em Data Science

Estatístico

- Aplicar teorias e métodos estatísticos para resolver problemas práticos de negócios, engenharia, ciências ou outras áreas de conhecimento
- Decidir quais dados são necessários para responder a perguntas ou problemas específicos
- Determinar métodos para encontrar ou coletar dados
- Realizar pesquisas de opinião para coletar dados
- Coletar dados ou treinar outras pessoas a fazê-lo
- Analisar e interpretar dados
- Relatar conclusões a partir de suas análises
- Decidir sobre uma estratégia adequada para coletar dados
- Extrair dados de fontes existentes ou instigar novos procedimentos (por exemplo, pesquisas com clientes, experiências científicas, sondagens de opinião, etc.)
- Analisar e interpretar dados usando ferramentas, algoritmos, modelos estatísticos e software (por exemplo R, SAS, SPSS, etc.)

Cursos recomendados na DSA: **Formação Cientista de Dados**

Cientista de Dados

www.datascienceacademy.com.br

Carreiras em Data Science

Cientista de Dados

- Comunicar previsões e resultados para a gestão e os departamentos de TI através de visualizações de dados eficazes
- Extrair grandes volumes de dados de múltiplas fontes internas e externas
- Empregar os programas de análise sofisticadas, aprendizado de máquina e métodos estatísticos para preparar os dados para uso em modelagem preditiva e prescritiva
- Explorar e analisar dados de uma variedade de ângulos para determinar fraquezas escondidas, tendências e / ou oportunidades
- Conceber soluções orientadas a dados para os principais desafios da empresa
- Criar novos algoritmos para resolver problemas e criar novas ferramentas para automatizar o trabalho
- Recomendar mudanças econômicas aos procedimentos e estratégias existentes

Cursos recomendados na DSA: **Formação Cientista de Dados, Formação Inteligência Artificial e Formação Java**

Funções Gerenciais em Data Science

Data Science Academy

Carreiras em Data Science

Funções Gerenciais em Data Science

Apesar de muitas denominações possíveis, 3 posições chave em liderança de projetos de Data Science são facilmente identificáveis:

- Líder de Equipes de Data Science
- Gerente de Dados e Analytics
- Chief Data Officer

Carreiras em Data Science

Funções Gerenciais em Data Science

Embora o volume de dados aumente a cada dia, o que requer investimento em armazenamento e análise, a má gestão dos dados ainda tem sido o que mais se vê no ambiente corporativo.

Gerenciar dados custa caro e de acordo com o Gartner, estima-se um prejuízo na ordem de 13 bilhões de dólares com o gerenciamento ineficaz dos dados.

Carreiras em Data Science

Funções Gerenciais em Data Science

Como os dados são armazenados eletronicamente, muitas empresas deixam sua gestão para o departamento de Tecnologia da Informação (TI).

Carreiras em Data Science

Funções Gerenciais em Data Science

No entanto, não há uma função em TI que seja voltada para o gerenciamento de dados e ninguém é oficialmente responsável pelos dados.

Carreiras em Data Science

Funções Gerenciais em Data Science

As pessoas esperam que os dados sejam armazenados com precisão, mas na maioria das vezes não é o que ocorre.

Carreiras em Data Science

Funções Gerenciais em Data Science

Entra em cena então o Chief Data Officer – CDO (Executivo Chefe de Dados)

Esta função é relativamente nova e muitas empresas ainda não estão dando atenção a isso

Carreiras em Data Science

Funções Gerenciais em Data Science

O papel do CDO é trazer ordem para o caos nos dados e proteger o investimento da empresa, seja na coleta, armazenamento ou análise de dados.

Carreiras em Data Science

Funções Gerenciais em Data Science

Tão ruim quanto armazenar dados de forma imprecisa, é analisar estes dados armazenados de forma imprecisa, o que pode levar a decisões de negócio catastróficas.

Como se Tornar um Cientista de Dados

Como se Tornar um Cientista de Dados?

É um momento emocionante para a Ciência de Dados.

Como se Tornar um Cientista de Dados?

O campo é novo, mas está crescendo rapidamente.

Há uma grande procura pelos Cientistas de Dados e a remuneração média nos EUA já é superior a 100 mil dólares por ano.

Como se Tornar um Cientista de Dados?

O déficit de competências de Ciência de Dados significa que muitas pessoas estão aprendendo ou tentando aprender Data Science.

Como se Tornar um Cientista de Dados?

O primeiro passo para aprender Ciência de Dados é geralmente perguntando ”Por onde começar?"

Como se Tornar um Cientista de Dados?

A resposta a esta pergunta tende a ser uma longa lista de cursos e livros para ler, começando com Álgebra Linear ou Estatística. É semelhante a um professor entregando uma pilha de livros ao aluno e dizendo ”leia tudo isso”.

Esse caminho tende a ser pouco motivador e na maiora das vezes, fracassa.

Como se Tornar um Cientista de Dados?

Algumas pessoas aprendem melhor com livros, outras com vídeos, mas a maneira mais eficaz de aprender, é fazendo

Como se Tornar um Cientista de Dados?

A motivação aparece quando você começa a ver o resultado do seu aprendizado

Como se Tornar um Cientista de Dados?

Como se Tornar um Cientista de Dados?

Se você quer aprender Ciência de Dados, seu primeiro objetivo deve ser aprender a estudar.

Como se Tornar um Cientista de Dados?

Interessado em saber como? Você está no lugar certo!

www.datascienceacademy.com.br

Data Science Academy

Como se Tornar um Cientista de Dados?

Quem é o Cientista de Dados?

"Data scientist is
the sexiest job
of the 21st century."

Harvard Business Review

A profissão mais sexy do século XXI!!

Como se Tornar um Cientista de Dados?

Quem é o Cientista de Dados?

"Um Cientista de Dados é alguém que é curioso, que analisa os dados para detectar tendências”.

Como se Tornar um Cientista de Dados?

Quem é o Cientista de Dados?

Cientistas de dados são uma nova geração de especialistas analíticos que têm as habilidades técnicas para resolver problemas complexos.

Como se Tornar um Cientista de Dados?

Quem é o Cientista de Dados?

E a curiosidade de explorar quais são os problemas que precisam ser resolvidos.

Como se Tornar um Cientista de Dados?

Quem é o Cientista de Dados?

Eles são parte matemáticos,
parte cientistas da computação e
parte analistas de tendências.

Como se Tornar um Cientista de Dados?

Quem é o Cientista de Dados?

E, por transitarem entre o mundo dos negócios e de TI, eles são muito procurados e bem pagos.

Como se Tornar um Cientista de Dados?

Quem é o Cientista de Dados?

Eles também são um sinal dos tempos modernos. Cientistas de dados não estavam no radar há uma década, mas sua popularidade repentina reflete como as empresas agora pensam sobre Big Data.

Como se Tornar um Cientista de Dados?

Quem é o Cientista de Dados?

Essa imensidão de dados não estruturados já não pode mais ser ignorada e esquecida. É uma mina de ouro virtual que ajuda a aumentar receitas - desde que haja alguém que escave e desenterre insights empresariais que ninguém havia pensado em procurar.

Como se Tornar um Cientista de Dados?

Entra em cena, o Cientista de Dados!

Como se Tornar um Cientista de Dados?

Quem é o Cientista de Dados?

Cientista de Dados

Pessoa que é melhor em estatística que um engenheiro de software e que é melhor em engenharia de software, que qualquer estatístico.

Como se Tornar um Cientista de Dados?

Quem é o Cientista de Dados?

- Profissional que pratica Ciência de Dados.
- Experiência em análise de dados, estatística, matemática e programação.
- Investiga problemas complexos de negócio e provê soluções a partir dos dados.

Como se Tornar um Cientista de Dados?

Características do Cientista de Dados

- Curiosidade
- Habilidade Técnica
- Comunicação
- Capacidade de apresentação
- Criatividade
- Conhecimento de Negócios

Como se Tornar um Cientista de Dados?

Características de um Cientista de Dados

- Contador de História
- Conhecimento de Negócio
- Machine Learning e Estatística
- Aplicação de Métodos Científicos
- Programação
- Curiosidade
- Engenharia de Dados

Como se Tornar um Cientista de Dados?

Desenvolvimento de Produtos	Dados não estruturados	Otimização	Programação Back e Front-end	Estatísticas temporais
Áreas de Negócio	Dados estruturados	Modelos gráficos		Estatística espacial
	Machine Learning	Algoritmos		Manipulação de dados
	Big Data e dados distribuídos	Simulações		Estatística clássica

Como se Tornar um Cientista de Dados?

Características do Cientista de Dados

É preciso profundo conhecimento em pelo menos uma das áreas abaixo:

- Estatística / Matemática
- Machine Learning
- Armazenamento / Processamento de Big Data
- Business

Como se Tornar um Cientista de Dados?

O que faz um Cientista de Dados?

Como se Tornar um Cientista de Dados?

Como se Preparar para a Profissão do Futuro?

Como se Tornar um Cientista de Dados?

Como se Tornar um Cientista de Dados?

O que o Cientista de Dados precisa saber?

Área de Conhecimento	Habilidade
Matemática e Estatística	Álgebra Linear, Estatística Descritiva, Testes de Hipótese, Análise Bayesiana
Aprendizado de Máquina	Aprendizagem Supervisionada e Não-Supervisionada, Classificação, Regressão, Clustering
Programação	Python, R, Scala, Java, Julia, SAS, SQL, C++
Banco de Dados	Bancos Relacionais e Bancos No-SQL como MongoDB
Filtragem e Visualização de Dados	D3.js, Tableau, Infovis, ggplot2
Big Data	Hadoop, Spark, Storm, Cassandra
Área de Negócio	Finanças, Marketing, Varejo, Astronomia, Saúde, Tecnologia

Como se Tornar um Cientista de Dados?

Em resumo...

Como se Tornar um Cientista de Dados?

Em resumo...

Como se Tornar um Cientista de Dados?

Por isso o Cientista de Dados é conhecido como Unicórnio.

Como se Tornar um Cientista de Dados?

Mas o Unicórnio é um mito

Se você quer participar desse mercado promissor, de Data Science e Big Data, foque em aprender o máximo possível e tenha em mente que o aprendizado é algo contínuo!

Como se Tornar um Cientista de Dados?

Mas o Unicórnio é um mito

Aprenda estatística de forma correta. Não adianta nada conseguir manipular bases de dados das maneiras mais mirabolantes possíveis, se o que você faz é somente totalizar as variáveis de dentro do seu banco de dados.

Como se Tornar um Cientista de Dados?

Mas o Unicórnio é um mito

Não tente aprender tudo de uma vez. Mantenha-se informado sobre as tecnologias, mas não se cobre demais por isso!

Como se Tornar um Cientista de Dados?

Seja um profissional de Analytics

Como se Tornar um Cientista de Dados?

Como se Tornar um Cientista de Dados?

1- Aprenda a amar os dados
2- Aprenda fazendo
3- Aprenda a comunicar seus insights
4- Aprenda com outros pares na comunidade
5- Aumente o grau de dificuldade e se desafie o tempo todo

Como se Tornar um Cientista de Dados?

Atualmente, não só o mercado depara-se com a escassez de profissionais bem capacitados, como aqueles que exercem a profissão devem estar preparados para os desafios que estão por vir.

Como se Tornar um Cientista de Dados?

Por certo, os Cientistas de Dados vieram para ficar.

Como se Tornar um Cientista de Dados?

Como se preparar para esta profissão, que é considerada a profissão do futuro?

Como se Tornar um Cientista de Dados?

www.datascienceacademy.com.br

Você Cientista De Dados

Esse é o nosso objetivo!

Muito Obrigado Por Acompanhar Este Treinamento!
Equipe Data Science Academy

Encerramento

Encerramento

É um momento emocionante para a Ciência de Dados

Encerramento

www.datascienceacademy.com.br

Estamos aqui para ajudá-lo!

Encerramento

E sua opinião é muito importante para nós! Mande suas críticas e sugestões!

Você tem um canal direto conosco:
contato@datascienceacademy.com.br

Encerramento

Lembre-se de fazer a avaliação final para obter seu certificado.

Muito Obrigado!

Continue Trilhando uma excelente jornada de aprendizagem.

Data Science

Big Data

Analytcis

Inteligência Artificial