

Guia Ágil

TechGuide - Alura

Inteligência Artificial

Nível 1

☐ Python para Ciência de Dados:

- Python é uma linguagem de programação interpretada de alto nível e que suporta múltiplos paradigmas de programação, como imperativo, orientado a objetos e funcional. É uma linguagem com tipagem dinâmica e gerenciamento automático de memória.
- Aprender lógica de programação em Python
- Aprenda os fundamentos da linguagem como variáveis, funções, listas, condicionais e imports
- Criar análises de dados
- Utilizar o Matplotlib para gerar gráficos
- Usar e manipular listas para agrupar dados
- Conhecer a biblioteca NumPy
- Conhecer a biblioteca Pandas

☐ Jupyter & Colab notebooks:

- Jupyter Notebook e Google Colaboratory são Notebooks que permitem a criação de blocos de texto e blocos de código
- Os Notebooks facilitam a elaboração de projetos de Data Science por ser possível visualizar o resultado da execução logo após o trecho de código

- O Google Colaboratory permite escrever e executar códigos Python diretamente no navegador, sem nenhuma ou poucas configurações necessárias
- Essas ferramentas facilitam o compartilhamento de projetos entre o time

☐ **Regressão Linear e Logística:**

- As regressões são os métodos mais simples de aprendizado supervisionado, porém encontram diversas aplicações.
- A regressão linear é usada para relacionar uma variável dependente contínua a uma ou mais variáveis independentes contínuas. O objetivo é encontrar uma relação linear que melhor se ajuste aos dados.
- A regressão logística, por outro lado, é usada para problemas de classificação binária, onde a variável de saída é categórica e possui apenas duas categorias. Ela estima a probabilidade de um evento ocorrer com base nas variáveis independentes.

☐ **Estatística e Matemática - Fundamentos:**

- Equações, funções e limites
- Logaritmos
- Matrizes, determinantes, vetores e espaço vetorial
- Derivadas e integrais
- Diferença entre média, mediana e moda
- Distribuição de frequência
- Variância e desvio padrão
- Distribuição binomial, poisson e normal
- Nível e intervalo de confiança
- Técnicas de amostragem
- Introdução à Regressão linear
- Séries temporais

☐ **Machine Learning - Fundamentos:**

- O Aprendizado de Máquina ou Machine Learning é um subcampo da Engenharia e da Ciência da Computação que evoluiu do estudo de reconhecimento de padrões e da teoria do aprendizado computacional em inteligência artificial.
- Aprendizado supervisionado
- Utilizar algoritmos de classificação
- Utilizar algoritmos de regressão
- Utilizar o Scikit-learn para criar modelos de machine learning

☐ **Machine Learning - Aprendizado Não Supervisionado:**

- O Aprendizado Não Supervisionado utiliza algoritmos de Aprendizado de Máquina para analisar e agrupar conjuntos de dados não rotulados. Estes algoritmos descobrem padrões ocultos ou agrupamentos de dados sem a necessidade de intervenção humana.
- O clustering ou análise de agrupamento de dados é o conjunto de técnicas de prospecção de dados (data mining) que visa fazer agrupamentos automáticos de dados segundo o seu grau de semelhança.
- Conhecer a análise exploratória
- Utilizar os métodos K-means, DBSCAN e Mean shift para agrupar dados sem classificação
- Avaliar a qualidade de uma clusterização
- Parametrizar métodos de clusterização através do máximo coeficiente de silhueta
- Entender a matemática por trás das métricas de validação Silhouette, Davies Bouldin e Calinski Harabasz
- Conhecer técnicas de redução de dimensionalidade

☐ **Naive Bayes:**

- Naive Bayes é um algoritmo de aprendizado de máquina comumente utilizado para classificação de texto e análise de sentimentos. Ele é baseado no teorema de Bayes, que descreve a probabilidade condicional de um evento ocorrer, dado que outro evento já ocorreu.

☐ Machine Learning - Máquina de Vetores de Suporte:

- Uma máquina de vetores de suporte (SVM, do inglês Support Vector Machine) é um conceito da Ciência da Computação para um conjunto de métodos de Aprendizado Supervisionado que analisam os dados e reconhecem padrões. O SVM pode ser usado tanto para modelar problemas preditivos de classificação quanto de regressão.

| Nível 2

☐ XGBoost:

- XGBoost é uma biblioteca de aprendizado de máquina baseada em gradient boosting que se destaca na resolução de problemas de classificação e regressão. "XG" em XGBoost refere-se a "Extreme Gradient", indicando o uso de gradient boosting para melhorar a precisão do modelo.

☐ Previsão de Séries temporais:

- A previsão de séries temporais (Time series Forecasting), também conhecida como previsão temporal, é uma técnica de análise estatística que envolve a previsão de valores futuros ou padrões com base em dados históricos ordenados no tempo. Em uma série temporal, os dados são coletados sequencialmente em intervalos regulares, como horários, diários, mensais ou anuais, e exibem dependências temporais.

☐ Deep Learning:

- Deep Learning é um ramo de Machine Learning baseado em um conjunto de algoritmos que tentam modelar abstrações de alto nível de dados usando um grafo profundo com várias camadas de processamento, compostas de várias transformações lineares e não lineares.
- Construir e treinar modelos com Keras
- Construir e treinar modelos com Tensorflow
- Selecionar as camadas de um modelo
- Classificar imagens
- Entender os conceitos de pesos e vieses

- Redes neurais para regressão
- Entender o conceito de Redes recorrentes

☐ **Visão Computacional:**

- Visão Computacional é um campo científico interdisciplinar que lida com a forma como os computadores podem ganhar conhecimentos de alto nível a partir de imagens ou vídeos digitais. Da perspectiva da engenharia, procura compreender e automatizar tarefas que o sistema visual humano pode fazer.
- Extrair regiões de interesse de uma imagem
- Normalizar e pré-processar dados de imagens
- Construir classificadores para reconhecimento de faces
- Extrair regiões do rosto humano baseado em marcos faciais
- Analisar diferentes condições de cada componente do rosto humano
- Conhecer Redes Neurais Convolucionais
- Usar OpenCV

☐ **Redes Neurais Convolucionais:**

- As CNNs (Convolutional Neural Networks, ou Redes Neurais Convolucionais) são um subtipo de redes neurais artificiais profundas (Deep Learning). AS CNNs são majoritariamente empregadas para o processamento de imagens e vídeos, no contexto de classificação do seu conteúdo. Ao contrário de outros algoritmos tradicionais de classificação de imagens, uma CNN tende a demandar menos esforços de pré-processamento nos dados de entrada.

☐ **Redução de Dimensionalidade:**

- Redução de dimensionalidade é uma técnica utilizada no campo de aprendizado de máquina e análise de dados para reduzir o número de variáveis ou dimensões em um conjunto de dados. O objetivo é simplificar a representação dos dados, eliminando ou combinando características redundantes ou irrelevantes, enquanto mantém a maior parte da informação original.

☐ **Machine Learning - Aprendizado por Transferência:**

- Transfer Learning (Aprendizado por Transferência) é uma abordagem na área de Machine Learning em que um modelo pré-treinado em uma tarefa relacionada é reutilizado como ponto de partida para treinar um modelo em uma tarefa diferente, geralmente com menos dados disponíveis.

☐ **Machine Learning - Regularização:**

- Regularização se refere a um conjunto de técnicas para tornar os dados mais regulares e estáveis, facilitando o aprendizado e a descoberta de padrões em Machine Learning. A regularização é uma técnica que reduz o erro de um modelo, evitando o sobreajustamento (Overfitting) e treinando o modelo para funcionar corretamente.

☐ **Chatbots:**

- Chatbots são agentes virtuais projetados para interagir com seres humanos por meio de uma interface de chat. Eles são baseados em técnicas de processamento de linguagem natural (NLP) e inteligência artificial (IA) para entender e responder às perguntas e comandos dos usuários de forma automatizada.

☐ **Segmentação Semântica e Instanciada de Imagens:**

- A segmentação semântica e instanciada de imagens é uma tarefa na área de visão computacional que envolve a atribuição de rótulos a pixels individuais em uma imagem. Ela tem como objetivo segmentar e identificar os objetos presentes na imagem, distinguindo diferentes classes de pixels e separando instâncias individuais dessas classes.

☐ **Visão Computacional - Extração de Características:**

- Feature Extraction (ou Extração de Características) é um processo na área de Machine Learning e Visão Computacional em que características relevantes e discriminantes são extraídas de um conjunto de dados para representar os objetos ou padrões de interesse. Essas características são transformadas em um novo conjunto de representações que possuem propriedades informativas para uma tarefa específica.

Nível 3

☐ **Processamento de Linguagem Natural:**

- Processamento de língua natural (PLN) é uma subárea da inteligência artificial e da linguística que estuda os problemas da geração e compreensão automática de línguas humanas naturais.
- Análise de Sentimento
- Criar visualizações para facilitar a análise de dados textuais
- Conhecer as bibliotecas NLTK e Scikit-Learn
- Normalizar textos
- Usar TF-IDF e Ngrams para melhorar a classificação
- Conhecer o conceito de Transformers e como são aplicados para LLMs
- Utilizar o SKlearn
- Utilizar Regex em PLN
- Conhecer o Word2Vec
- Combinar vetores de palavras para representar textos e classificá-los

☐ **Redes Neurais Recorrentes:**

- Redes Neurais Recorrentes (em inglês, Recurrent Neural Networks ou RNNs) são uma classe de redes neurais artificiais projetadas para lidar com dados sequenciais, como séries temporais, texto e fala. Ao contrário das redes neurais tradicionais, as RNNs têm conexões recorrentes que permitem que as informações sejam propagadas ao longo do tempo, permitindo que elas capturem dependências temporais em sequências de dados.
- LSTM (Long Short-Term Memory) é uma arquitetura de rede neural recorrente (RNN) projetada para superar as limitações das RNNs tradicionais na modelagem de dependências de longo prazo em sequências de dados.
- Conhecer a arquitetura de Redes Feedforward
- Funções de ativação mais comuns
- Familiarizar-se com as principais arquiteturas de RNNs, como bidirecionais (BRNN), Memória de curto prazo longa (LSTM), Unidades recorrentes

fechadas (GRU)

☐ **Redes Adversárias Generativas (GANs):**

- Redes Adversárias Generativas (em inglês, Generative Adversarial Networks ou GANs) são um tipo de arquitetura de rede neural usada para gerar dados sintéticos, como imagens, música, texto ou outros tipos de dados, que se assemelham aos dados de treinamento.

☐ **Aprendizado por Reforço:**

- O Aprendizado por Reforço é uma área de Machine Learning que se preocupa com a forma como agentes inteligentes devem tomar medidas num ambiente, a fim de maximizar a noção de recompensa cumulativa.
- Entender os conceitos de agente e recompensa
- Entender a diferença entre reforço positivo e negativo
- Conhecer o modelo Markov Decision Process
- Entender o conceito de Retorno
- Utilizar o algoritmo Q-learning

☐ **Viés Cognitivo e IA Responsável:**

- O conceito de Viés Cognitivo refere-se aos padrões sistemáticos de desvio e distorção na forma como os seres humanos percebem, interpretam e tomam decisões. Esses vieses cognitivos podem levar a distorções na análise de informações, julgamentos e comportamentos de todos os tipos, como de raça, gênero, idade, capacidades físicas, entre outros, influenciando nossas percepções e ações de maneira não racional ou imparcial.

☐ **Lógica Fuzzy:**

- A lógica fuzzy, também conhecida como lógica difusa, é uma extensão da lógica clássica que permite o tratamento de incerteza e imprecisão nos sistemas de tomada de decisão. Enquanto a lógica clássica trabalha com valores binários (verdadeiro ou falso), a lógica fuzzy lida com valores contínuos entre 0 e 1, representando graus de pertinência.

- Conhecer o conceito de valor crisp
- Como aplicar a fuzzificação
- Conhecer o método Mamdani

☐ **Automação com IA:**

- Automated IA, ou automação de IA, refere-se à combinação de tecnologias de automação, como RPA, com recursos de inteligência artificial, como processamento de linguagem natural (NLP), aprendizado de máquina (machine learning) e visão computacional. Essa abordagem visa automatizar processos complexos que envolvem o processamento e a análise de informações não estruturadas, permitindo que as máquinas tomem decisões com base nesses dados.

Habilidade Auxiliar: Business e Serviços

☐ **Big Data - Fundamentos:**

- Big Data refere-se a conjuntos de dados extremamente grandes e complexos, que não podem ser facilmente processados ou gerenciados por métodos tradicionais. Esses conjuntos de dados são caracterizados por seu volume massivo, velocidade de geração e variedade de tipos e formatos. Com a explosão da quantidade de dados gerados diariamente, provenientes de diversas fontes como redes sociais, dispositivos IoT e transações comerciais, o Big Data apresenta desafios e oportunidades para extrair insights valiosos e tomar decisões informadas.
- A análise de Big Data envolve o uso de técnicas e tecnologias avançadas, como armazenamento distribuído, processamento em paralelo e aprendizado de máquina, para explorar e transformar esses dados em informações significativas para empresas, organizações e pesquisadores.

☐ **Git e GitHub - Fundamentos:**

- Git é um sistema de controle de versão distribuído gratuito e de código aberto projetado para lidar com tudo, desde projetos pequenos a muito grandes com velocidade e eficiência.

- GitHub é um serviço de hospedagem para desenvolvimento de software e controle de versão usando Git.
- Criar um repositório
- Clonar um repositório
- Fazer commit, push e pull de e para o repositório
- Reverter um commit
- Criar branches e pull requests
- Lidar com merge e conflitos

☐ **Governança em IA e Business Analytics:**

- A governança em Business Analytics refere-se à gestão dos processos e estratégias de análise de dados nas organizações. Isso envolve a definição de políticas e procedimentos para a coleta, organização, análise e interpretação dos dados para obter insights valiosos e embasar a tomada de decisões empresariais.

☐ **Plataformas e Serviços Cognitivos:**

- Plataformas e serviços são projetados para fornecer recursos avançados de processamento de linguagem natural, reconhecimento de fala, visão computacional, análise de sentimentos e outras capacidades cognitivas.
- Como lidar com pipelines
- Conhecer o Databricks
- Conhecer o PySpark
- Conhecer o SageMaker

☐ **Engenharia de Dados - Fundamentos:**

- Uma pessoa Engenheira de Dados desempenha um papel crucial ao projetar, implementar e manter as infraestruturas de dados que permitem que uma organização trabalhe de maneira eficiente com suas informações. A pessoa engenheira de dados cria pipelines de dados para integrar, limpar e transformar dados provenientes de várias fontes e formatos. Isso

possibilita a geração de insights valiosos para o negócio e melhora a tomada de decisões estratégicas.

☐ **SQL - Fundamentos:**

- SQL (Structured Query Language, traduzindo, Linguagem de Consulta Estruturada) é uma linguagem de programação padronizada que é usada para gerenciar bancos de dados relacionais e realizar várias operações sobre os dados neles contidos.
- Conhecer os comandos mais comuns do SQL
- Usar SELECT para consultar uma tabela
- Usar INSERT para inserir dados em uma tabela
- Usar UPDATE para atualizar uma tabela
- Usar DELETE para remover dados de uma tabela
- Usar JOIN para conectar os dados de múltiplas tabelas
- Conhecer as cláusulas (FROM, ORDER BY, etc)

☐ **Cloud - Provedores:**

- Conhecer os principais provedores de serviços em nuvem e suas ofertas.
- Comparar as características e serviços oferecidos por diferentes provedores.
- Entender como selecionar o provedor de serviços em nuvem adequado para suas necessidades.
- Explorar as tendências e inovações no mercado de serviços em nuvem.

☐ **MLOps:**

- O MLOps permite que os modelos de Machine Learning sejam implantados de maneira rápida e confiável, o que é especialmente importante em empresas que lidam com grandes volumes de dados.
- Ajuda a garantir a qualidade e confiabilidade dos modelos de Machine Learning em produção, além de facilitar a manutenção e atualização desses modelos.

- Permite que as pessoas engenheiras de dados trabalhem em conjunto com cientistas de dados e desenvolvedores de software para implementar soluções de Machine Learning em larga escala.
- Ajuda a garantir a governança e a conformidade dos modelos de Machine Learning com as políticas e regulamentações da empresa.

☐ **Habilidades de comunicação:**

- Um bom nível de comunicação facilita o atingimento de objetivos, resolução de problemas, além de aumentar a produtividade, porque cada profissional saberá exatamente o que se espera dele e transmitir com clareza suas ideias.

☐ **Contêineres:**

- Os contêineres são pacotes de software que contêm todos os elementos necessários para serem executados em qualquer ambiente. Gerenciamento de contêineres é uma área crucial na computação em nuvem e DevOps, que envolve o uso de tecnologias para automatizar o processo de criação, implantação, escalonamento e monitoramento de contêineres. Contêineres são unidades de software padronizadas que permitem aos desenvolvedores empacotar todas as dependências de um aplicativo (código, bibliotecas, configurações, etc.) em um único pacote. Isso permite que o aplicativo seja executado de forma consistente em qualquer ambiente de infraestrutura.
- A tecnologia de contêineres, como exemplificada pelo Docker, fornece um ambiente consistente e portátil para desenvolvimento, teste e implantação de aplicativos, o que é vital para o trabalho eficiente de engenharia de dados. Além disso, o Kubernetes, um sistema de orquestração de contêineres, permite o gerenciamento, a automação e a escalabilidade de aplicações baseadas em contêineres em ambientes de produção. Dominar esses conceitos e tecnologias possibilita a engenheiros de dados construir e manter pipelines de dados eficientes e confiáveis.
- O Kubernetes (também conhecido como k8s ou kube) é uma plataforma de orquestração de containers open source que automatiza grande parte dos processos manuais necessários para implantar, gerenciar e escalar aplicações em containers.

- Isolar seu software para funcionar independentemente
- Implantar software em clusters
- Modularizar seu sistema em pacotes menores
- Conhecer a plataforma Docker
- Conhecer Kubernetes

Habilidade Auxiliar: Hardware e Eletrônica

☐ Robótica na IA - Fundamentos:

- Uma das vertentes da revolução 4.0 é a implementação em larga escala industrial da IA aos sistemas robotizados, podendo desenvolver processos autônomos e autogerenciáveis em campo.

☐ Hardwares para IA:

- Hardwares para IA são componentes físicos, como processadores, placas de vídeo e sistemas especializados, que são projetados e otimizados para realizar tarefas relacionadas à inteligência artificial de maneira eficiente e rápida. Eles são especialmente projetados para lidar com as demandas computacionais intensivas e algoritmos complexos usados em modelos de aprendizado de máquina e deep learning.

☐ Computação Paralela & Clusters:

- Computação paralela refere-se ao uso de múltiplos processadores ou unidades de processamento para executar tarefas computacionais simultaneamente, em paralelo. Em vez de um único processador executar uma tarefa de cada vez, a computação paralela divide a carga de trabalho em várias partes menores que podem ser processadas em paralelo, resultando em maior eficiência e velocidade de processamento.

☐ Internet das Coisas (IoT) - Fundamentos:

- Quase todos os objetos agora podem ser conectados à Internet, de máquinas de lavar a componentes sofisticados de motor de avião, a itens orgânicos, como plantações e vacas. A capacidade de conectar coisas e

capturar dados úteis está transformando as empresas de todos os setores e abrindo portas para novas especializações de carreira. Descubra a base desse novo campo com atividades práticas para conectar os sensores com segurança aos serviços em nuvem nas redes IP e coletar dados em um sistema de IoT (Internet das Coisas) de ponta a ponta.

- Como rodar seu primeiro sistema em um Arduino

☐ **Wearables e Aquisição de Dados:**

- Wearables é a expressão que representa as tecnologias vestíveis que além de captar biopotenciais podem gerar dados de grande valia, como relógios e óculos inteligentes, e também na área na saúde para pessoas com alguma deficiência física.