|  |
| --- |
| **CONTEÚDO TEÓRICO – AULA 02**  **DISCIPLINA: Data Science, Marchine Learning, Data Minning** |
| **Data Science, Machine Learning e Data Mining: Uma Visão Geral**  Nos últimos anos, termos como Data Science, Machine Learning e Data Mining tornaram-se cada vez mais proeminentes no mundo da tecnologia e dos negócios. Esses campos estão na vanguarda da revolução digital, desempenhando um papel crucial na análise e interpretação de grandes volumes de dados para obter insights valiosos e impulsionar a tomada de decisões informadas. Neste documento, exploraremos em detalhes o que é Data Science, Machine Learning e Data Mining, suas aplicações e como eles se relacionam entre si.  **Data Science**  Data Science é um campo multidisciplinar que combina conhecimentos de estatística, programação, ciência da computação e domínio do assunto para extrair conhecimentos e insights dos dados. O objetivo principal da Data Science é transformar dados brutos em informações úteis e acionáveis que possam orientar a tomada de decisões em uma variedade de áreas, incluindo negócios, saúde, finanças, entre outros.  Componentes da Data Science:   * Coleta de Dados: O primeiro passo na análise de dados é a coleta de dados relevantes de várias fontes, como bancos de dados, sensores, mídias sociais, entre outros. * Limpeza e Preparação de Dados: Os dados coletados geralmente contêm erros, valores ausentes ou inconsistentes, portanto, é necessário limpar e preparar os dados para análise, removendo ruídos e garantindo sua qualidade. * Análise Exploratória de Dados (EDA): A EDA envolve a exploração dos dados por meio de técnicas estatísticas e visualizações para identificar padrões, tendências e relações entre as variáveis. * Modelagem de Dados: Nesta fase, modelos estatísticos e algoritmos de Machine Learning são aplicados aos dados para fazer previsões, classificações ou identificar padrões ocultos. * Interpretação e Comunicação de Resultados: Por fim, os resultados da análise são interpretados e comunicados de forma clara e eficaz para orientar a tomada de decisões.   **Machine Learning**  Machine Learning é uma subárea da inteligência artificial (IA) que se concentra no desenvolvimento de algoritmos e técnicas que permitem aos computadores aprenderem a partir dos dados e fazer previsões ou tomar decisões sem serem explicitamente programados. Em vez de seguir instruções específicas, os sistemas de Machine Learning são treinados com exemplos de dados para reconhecer padrões e fazer inferências.  Tipos de Aprendizado de Máquina:   * Aprendizado Supervisionado: Nesse tipo de aprendizado, os algoritmos são treinados com exemplos rotulados, ou seja, o modelo aprende a fazer previsões com base em pares de entrada e saída fornecidos. * Aprendizado Não Supervisionado: Aqui, o modelo é treinado com exemplos não rotulados, e o objetivo é encontrar padrões ou estrutura nos dados sem a necessidade de rótulos predefinidos. * Aprendizado por Reforço: Neste tipo de aprendizado, o modelo aprende através de tentativa e erro, recebendo feedback positivo ou negativo com base em suas ações.   Aplicações de Machine Learning:   * Reconhecimento de padrões em imagens e vídeos. * Análise de sentimentos em redes sociais. * Recomendação de produtos em plataformas de comércio eletrônico. * Diagnóstico médico assistido por computador. * Previsão de demanda em cadeias de suprimentos.   **Data Mining**  Data Mining é o processo de descoberta de padrões e relações significativas em grandes conjuntos de dados. Ele utiliza técnicas de análise estatística e de Machine Learning para explorar grandes volumes de dados em busca de insights valiosos que possam ser usados para tomada de decisões estratégicas.  Técnicas de Data Mining:   * Associação: Identifica padrões de associação entre diferentes variáveis em um conjunto de dados, como a associação de itens frequentemente comprados juntos em transações de varejo. * Classificação: Classifica os dados em categorias ou classes com base em características comuns, como classificar emails como spam ou não spam. * Clustering: Agrupa dados semelhantes em clusters ou grupos com base em suas características, permitindo a identificação de padrões intrínsecos nos dados. * Previsão: Usa modelos estatísticos ou de Machine Learning para fazer previsões sobre eventos futuros com base em dados históricos.   Diagrama  Descrição gerada automaticamente |