

TECNOLOGIAS EMERGENTES: MINERAÇÃO DE DADOS

JOSÉ RAMON DA CONCEIÇÃO MONTEIRO
RENATO HIDAKA TORRES

2024

FICHA TÉCNICA

Realização

Universidade Federal do Pará
 Instituto de Filosofia e Ciências Humanas
 Programa de Pós-graduação em Segurança Pública

Apoio

Escola Estadual de Ensino Médio Agroindustrial Juscelino Kubitschek de Oliveira

Supervisão

Renato Hidaka Torres

Elaboração e Organização

José Ramon da Conceição Monteiro

Ilustração e Design

José Ramon da Conceição Monteiro

Como referenciar esta obra

MONTEIRO, José Ramon da Conceição; TORRES, Renato Hidaka. **Tecnologias Emergentes: Mineração de Dados**. Programa de Pós-graduação em Segurança Pública. Instituto de Filosofia e Ciências Humanas. Universidade Federal do Pará. Belém, Pará, Brasil, 2024.

Agradecimentos

À Escola Estadual de Ensino Médio Agroindustrial Juscelino Kubitschek de Oliveira, ao Programa de Pós-Graduação em Segurança Pública (PPGSP), do Instituto de Filosofia e Ciências Humanas (IFCH), da Universidade Federal do Pará (UFPA), à Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) e ao Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq).



Sumário

Apresentação	04
Mineração de Dados	05
Principais Algoritmos	06
Como escolher o algoritmo certo para um problema específico?	07
Quais são as principais métricas para avaliar o desempenho de um algoritmo?	08
Conclusão	09
Referências	10

Apresentação



Esta cartilha foi elaborada a partir dos estudos realizados no Mestrado do Programa de Pós-Graduação em Segurança Pública da UFPA, cuja temática de estudo foi “Mineração de Dados e Explicabilidade de Aprendizado de Máquina Aplicadas à Segurança Pública”. Tecnologias emergentes são inovações tecnológicas que têm o potencial para criar ou transformar o ambiente de negócios nos próximos 5 a 10 anos e que poderão alcançar grande influência econômica, mas que ainda não se consolidaram. Essas inovações técnicas ou em desenvolvimento têm características que as tornam promissoras, mas ainda não fazem parte do cotidiano das pessoas. Elas estão em fases experimentais, de estudo ou em fase de amadurecimento, sendo usadas por empresas inovadoras.

Essas tecnologias emergentes podem abranger diversas áreas e categorias, desde a inteligência artificial até a mineração de dados e o aprendizado de máquina. Elas estão em constante evolução e, uma vez utilizadas pela população, conseguem influenciar significativamente como as pessoas trabalham, se relacionam, consomem ou seguem com suas rotinas. O termo “emergentes” é usado porque esses recursos ainda não foram devidamente consolidados.

O objetivo desta cartilha, é apresentar os principais conceitos relacionados às técnicas de Mineração de Dados, os principais algoritmos utilizados, como escolher o algoritmo mais adequado para resolver problemas específicos e quais métricas utilizar para avaliar o desempenho do algoritmo.

Mineração de Dados



Mineração de Dados, também conhecida como **Descoberta de Conhecimento em Dados (KDD)**, é um processo técnico que envolve a análise de grandes conjuntos de dados para descobrir padrões, correlações e outras informações valiosas. O objetivo é transformar dados brutos em conhecimento útil, que pode ser utilizado para tomada de decisões, previsão de tendências e melhor entendimento de fenômenos complexos.

O processo de Mineração de Dados geralmente consiste em quatro etapas principais:

1

Definição dos objetivos do negócio: Esta etapa envolve a identificação clara do problema a ser resolvido ou da questão a ser respondida. É crucial para direcionar o processo de mineração de forma eficaz.

2

Preparação dos dados: Os dados são coletados, limpos e transformados para garantir que sejam adequados para análise. Esta etapa é fundamental, pois a qualidade dos dados influencia diretamente a qualidade dos insights obtidos.

3

Aplicação de algoritmos de mineração de dados: Utilizam-se métodos estatísticos, algoritmos de aprendizado de máquina e técnicas de visualização para explorar os dados. Estes métodos ajudam a identificar padrões e relações que não seriam visíveis de outra forma.

4

Avaliação dos resultados: Os padrões e informações descobertos são avaliados para determinar sua utilidade e relevância para os objetivos definidos.

A Mineração de Dados é utilizada em diversos setores, como varejo, finanças e telecomunicações, para descobrir relações entre variáveis, detectar fraudes, entender o comportamento do consumidor, entre outros

usos. Com o crescimento do Big Data, a Mineração de Dados tornou-se ainda mais importante, pois permite que as organizações extraiam insights valiosos de volumes massivos de dados.

Principais Algoritmos



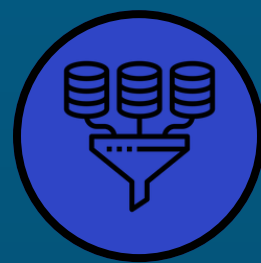
Os principais algoritmos de mineração de dados são fundamentais para extrair informações valiosas de grandes conjuntos de dados. Aqui estão alguns dos mais utilizados:



Árvores de Decisão: Utilizam uma estrutura de árvore para representar decisões e suas possíveis consequências, incluindo chance de evento, custos e utilidade.



K-Means: É um método de clusterização que divide um conjunto de dados em k grupos distintos, sendo k um número pré-definido.



Regras de Associação: Buscam identificar regras que descrevem fortemente as relações entre itens em um conjunto de dados.



Naive Bayes: Um classificador probabilístico baseado no teorema de Bayes, útil em situações onde os atributos são independentes.



Redes Neurais: Modelos computacionais inspirados pelo sistema nervoso humano, capazes de reconhecer padrões complexos.



Máquinas de Vetores de Suporte (SVM): Algoritmos de aprendizado supervisionado usados para classificação e regressão.

Esses algoritmos são aplicados de acordo com o tipo de dados e o objetivo da análise, permitindo que os cientistas de dados descubram padrões, tendências e associações que não seriam visíveis de outra forma.



Como escolher o algoritmo certo para um problema específico?

Escolher o algoritmo de mineração de dados correto para um problema específico é uma decisão importante que pode impactar significativamente a eficácia da análise. Aqui estão alguns passos e considerações que podem ajudar nesse processo:

1

Defina o Tipo de Problema: Identifique se o problema é de classificação, regressão ou agrupamento. Cada tipo de problema tem algoritmos mais adequados.

2

Tamanho e Qualidade dos Dados: Considere o volume de dados disponíveis e sua qualidade. Algoritmos diferentes têm desempenhos variados dependendo do tamanho do conjunto de dados.

3

Complexidade dos Dados: Avalie a complexidade dos dados. Alguns algoritmos lidam melhor com dados não lineares ou com muitas variáveis.

4

Interpretabilidade dos Resultados: Pense sobre a necessidade de explicar os resultados. Modelos como árvores de decisão são mais fáceis de interpretar do que redes neurais, por exemplo.

5

Recursos Computacionais: Leve em conta os recursos computacionais disponíveis. Algoritmos mais complexos podem exigir mais poder de processamento.

6

Experiência e Conhecimento: Utilize sua experiência e conhecimento prévio. Algoritmos com os quais você está familiarizado podem ser mais fáceis de ajustar e otimizar.

Além destes passos, é útil realizar testes com diferentes algoritmos e com-

parar seus resultados. A escolha do algoritmo pode ser um processo iterativo, onde ajustes são feitos com base no desempenho e nos resultados obtidos. Consultar especialistas em mineração de dados ou utilizar ferramentas de software que sugerem algoritmos também pode ser benéfico.

Quais são as principais métricas para avaliar o desempenho de um algoritmo?

As principais métricas para avaliar o desempenho de um algoritmo de mineração de dados são:



Taxa de Acerto (Accuracy): Mede a precisão do modelo, representando a porcentagem de previsões corretas feitas pelo algoritmo.



Tempo de Execução: Avalia a eficiência do algoritmo em termos de velocidade de processamento.



Taxa de Descoberta de Padrões: Indica a capacidade do algoritmo de encontrar informações relevantes e insights nos dados.



Taxa de Falsos Positivos e Falsos Negativos: Mede a qualidade das previsões, onde valores baixos indicam um modelo mais preciso.



Escalabilidade: Refere-se à capacidade do algoritmo de lidar com o aumento no volume de dados.



Taxa de Cobertura: Mostra a capacidade do modelo de abranger todas as possíveis situações presentes nos dados.

Essas métricas ajudam a determinar a adequação de um algoritmo para um determinado conjunto de dados e objetivo de análise, contribuindo para a escolha da melhor abordagem em projetos de mineração de dados.

Conclusão



A Mineração de Dados é um campo vital da ciência da computação e da análise de dados que se concentra na extração de informações significativas de grandes conjuntos de dados. Ela permite que as organizações descubram padrões ocultos, tendências e correlações, transformando dados brutos em insights acionáveis para a tomada de decisões estratégicas.

Os principais algoritmos de mineração de dados incluem Árvores de Decisão, K-Means, Regras de Associação, Naive Bayes, Redes Neurais e Máquinas de Vetores de Suporte (SVM). Cada um desses algoritmos tem suas forças e é adequado para diferentes tipos de problemas de dados. Para escolher o melhor algoritmo, é essencial entender o problema em mãos, conhecer os dados disponíveis e considerar fatores como a complexidade do modelo, a facilidade de interpretação, os recursos computacionais disponíveis e a experiência prévia com algoritmos específicos. A escolha do algoritmo certo é muitas vezes um processo iterativo que envolve experimentação e ajuste fino.

As métricas de avaliação são cruciais para medir o desempenho dos algoritmos de mineração de dados. Essas métricas ajudam a determinar a eficácia de um algoritmo e garantir que ele atenda aos objetivos do projeto.

Em resumo, a mineração de dados é um campo dinâmico e em constante evolução que desempenha um papel crucial na era do Big Data. A habilidade de escolher o algoritmo certo e avaliar seu desempenho adequadamente é fundamental para extrair o máximo valor dos dados disponíveis.

Referências



CAÑADAS, R. Algoritmos de mineração de dados. **ABDATUM**, 11 mar. 2024. Disponível em: [<https://abdatum.com/pt/tecnologia/algoritmos-de-mineracao-de-dados>]. Acesso em: 09 maio 2024.

DATA MINING: definição, exemplos e aplicações - Descubra como o 'data mining' irá prever nosso comportamento. **IBERDROLA**, [S. D.], Inovação. Disponível em: [<https://www.iberdrola.com/inovacao/data-mining-definicao-exemplos-e-aplicacoes>]. Acesso em: 09 maio 2024.

DESEMPENHO NA MINERAÇÃO DE DADOS: otimizando resultados e eficiência. **AWARI**, 30 jul. 2023, Ciências de Dados. Disponível em: [<https://awari.com.br/desempenho-na-mineracao-de-dados-otimizando-resultados-e-eficiencia/>]. Acesso em: 09 maio 2024.

DIAS, M. M. **Parâmetros na escolha de técnicas e ferramentas de mineração de dados**. Acta Scientiarum, Maringá : São Paulo, v. 24, n. 6, p. 1715-1725, 2002.

INTRODUÇÃO À MINERAÇÃO DE DADOS: Algoritmos e técnicas essenciais. **AWARI**, 30 jul. 2023, Ciências de Dados. Disponível em: [<https://awari.com.br/introducao-a-mineracao-de-dados-algoritmos-e-tecnicas-essenciais/>]. Acesso em: 09 maio 2024.

MINERAÇÃO DE DADOS: O que é e qual sua importância?. **SAS**, [s. d.]. Disponível em: [www.sas.com/pt_br/insights/analytics/mineracao-de-dados.htmlr-pt/topics/data-mining]. Acesso em: 09 maio 2024.

Referências



NASCIMENTO, J. C. D. Mineração de Dados: definição, objetivos, vantagens e modelagem. **Instituto Eldorado**, 26 fev. 2021, Data Science. Disponível em: [<https://www.eldorado.org.br/blog/mineracao-de-dados-definicao-objetivos-vantagens-e-modelagem/>]. Acesso em: 09 maio 2024.

O que é e quais são as principais tecnologias emergentes? Entenda aqui. **BLOG ALOO**, 02 mar. 2021, Tendências. Disponível em: [<https://blog.aloo.com.br/tecnologias-emergentes/>]. Acesso em: 09 maio 2024.

O que é mineração de dados?. **IBM**, [s. d.]. Disponível em: [<https://www.ibm.com/br-pt/topics/data-mining>]. Acesso em: 09 maio 2024.

O que é tecnologia emergente e quais os principais exemplos. **AMCHAM**, 11 jan. 2024. Disponível em: [<https://www.amcham.com.br/blog/tecnologia-emergente>]. Acesso em: 09 maio 2024.

O que são tecnologias emergentes e quais as principais. **SANKHYA**, 19 jul. 2023, Tecnologia e Tendências. Disponível em: [<https://www.sankhya.com.br/blog/tecnologias-emergentes/>]. Acesso em: 09 maio 2024.

TECNOLOGIAS EMERGENTES: MINERAÇÃO DE DADOS

