

UNIVERSIDADE DE BRASÍLIA
Faculdade do Gama

Sistemas de Banco de Dados 2

Trabalho Final (TF)

Data Mining (Mineração de Dados)

Adne Moreira - 200013181

Brasília, DF

2023

Definição da Tecnologia Pesquisada

De acordo com Berry e Linoff no ano de 1997, mineração de dados ou *Data Mining* é a exploração e a análise de grandes quantidades de dados, de forma automática com o intuito de descobrir padrões e regras, utilizam de modelos de inteligência artificial. Tal técnica de exploração de dados foi proposta na década de 80.

Dessa forma, é possível entender que mineração de dados é muito útil para identificar padrões e coletar informações que podem auxiliar a resolver e prever problemas de negócio e análise de dados, auxiliando na tomada de decisões. Mineração de dados é uma parte do Knowledge Discovery in Database (KDD) ou Descoberta de Conhecimento em Bases de Dados.

KDD é uma abordagem sistemática de descobrir conhecimentos úteis em um conjunto de dados. Engloba diversas etapas, desde a seleção e preparação de dados até a interpretação dos seus resultados. Para o entendimento da etapa principal do KDD, que é a mineração de dados, é importante entender as etapas que antecedem, sendo assim, as etapas do KDD são:

- Seleção de dados: Os dados relevantes para o contexto específico são selecionados;
- Pré-processamento: Os dados passam por um processo de limpeza e de transformação, para tratamento de ruídos;
- Transformação: Os dados são transformados em um formato específico para análise;
- Mineração de Dados: É a etapa em que são utilizados algoritmos de aprendizado de máquina e técnicas de mineração de dados para descobrir padrões e tendências;
- Avaliação: Os dados que foram obtidos pela mineração de dados são analisados quanto a relevância, utilidade para o contexto;
- Interpretação e visualização: Dados interpretados e visualizados de maneira a facilitar a visualização.

A imagem 1 representa cada etapa da Descoberta de Conhecimento em Bases de Dados.

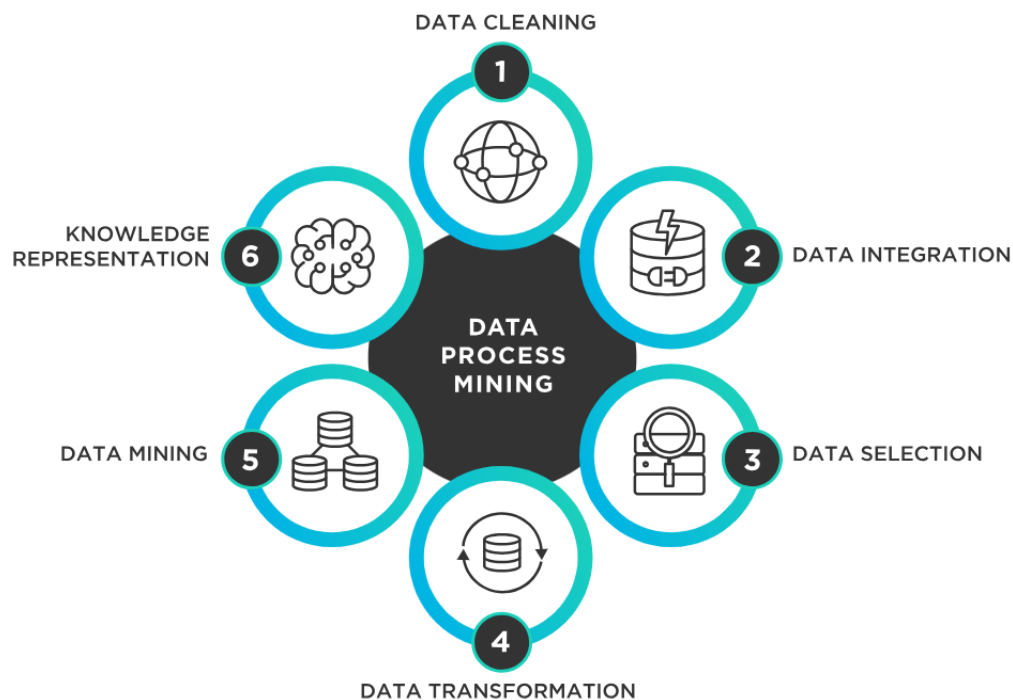


Figura 1 – Data Mining

Fonte: <https://www.tibco.com/pt-br/reference-center/what-is-data-mining>

A mineração de dados surgiu como necessidade para transformação de dados que muitas vezes eram inúteis em conhecimento que pode ser utilizado para estratégias e tomada de decisões em organizações, pois muitas vezes muito era gasto no armazenamento de dados que eram pobres em informação, pois dado não é o mesmo de informação.

Dado é a representação bruta de fatos, números que não possuem um contexto ou um significado. Já informação é o dado organizado e contextualizado de forma a dar sentido ao dado, transformação do dado em algo que pode ser compreendido. Já conhecimento consiste em uma interpretação mais profunda da informação para tomada de decisão em algum contexto.

A utilização de Data Mining pode ser de suma importância para diversas áreas, incluindo Marketing, CyberSegurança, detecção de fraude, gerenciamento de riscos, etc. Pelo seu potencial, a mineração de dados é crucial para iniciativas analíticas em organizações e auxilia em diversos aspectos de negócios.

Existem algumas técnicas que podem ser utilizadas para mineração de

dados em diferentes aplicações de ciência de dados. Algumas das técnicas mais conhecidas e utilizadas são:

- **Mineração de regras de associação:** Na mineração de dados, regras de associação são condições de if-then que identifica relacionamentos entre elementos de dados. O suporte mede o número de vezes com que os elementos relacionados aparecem em um conjunto de dados.
- **Classificação:** Essa abordagem na mineração de dados atribui os elementos em conjuntos de dados, classificando-os em diferentes categorias.
- **Agrupamento:** No agrupamento, os elementos de dados que compartilham algumas características específicas são agrupados em clusters;
- **Regressão:** Se trata de uma outra forma de encontrar relacionamentos entre os elementos de um conjunto de dados, calculando valores de dados previstos com base em um conjunto de variáveis.

As ferramentas de mineração de dados estão presentes em diversos fornecedores e incluem recursos de preparação de dados, algoritmos integrados e suporte a uma abordagem preditiva.

Sendo assim, a técnica de mineração de dados pode levar a uma melhoria significativa da confiabilidade na tomada de decisões.

Objetivos da tecnologia pesquisada

A mineração de dados tem como principal objetivo identificar padrões, relações e informações valiosas em conjuntos de dados extensos. Essa técnica possibilita a descoberta de conhecimentos relevantes e o aproveitamento de insights que não seriam facilmente perceptíveis de outra forma.

Além disso, a utilização da mineração de dados apresenta uma série de objetivos específicos. Entre eles:

- **Descoberta de conhecimento:** Busca identificar conhecimentos úteis que não são facilmente perceptíveis;
- **Previsão e tomada de decisões:** A mineração de dados pode ser

utilizada para desenvolver modelos preditivos, para prever com base em padrões identificados e elaborar estratégias;

- Segmentação de dados: A mineração de dados também pode ser utilizada para segmentar dados em diferentes grupos a partir de características em comum com outros dados;
- Otimização de processos: Pode otimizar processos de negócios em organizações e oportunidades de melhorias na eficiência operacional;
- Detecção de anomalias ou melhorias: Buscando identificar anomalias e oportunidades de melhorias em dados de diversas áreas, como cybersegurança e detecção de fraudes. Auxiliando na melhoria da segurança de organizações.

Vantagens da tecnologia pesquisada

A técnica de mineração de dados tem sido amplamente utilizada e pode trazer diversas vantagens significativas, algumas das vantagens incluem:

- Tomada de decisões baseadas em dados: A mineração de dados auxilia na tomada de decisões baseadas na análise de dados e de fatos já conhecidos;
- Previsão de tendência e comportamentos: Permite desenvolver modelos preditivos de tendência e comportamentos;
- Descoberta de conhecimentos: Permite descobrir conhecimentos valiosos em um conjunto de dados, relevando padrões complexos;
- Vantagem competitiva: A utilização de mineração de dados pode ser uma vantagem competitiva das organizações, pois possibilita um entendimento mais aprofundado dos clientes, mercado e tendências, sendo essencial para estratégias.

Sendo assim, a mineração de dados é uma técnica poderosa para a extração de valor e conhecimento dos dados.

Desvantagens da tecnologia pesquisada

Assim como toda técnica e tecnologia, a mineração de dados apresenta algumas desvantagens em sua utilização. É importante conhecer as desvantagens e vantagens da tecnologia para a aplicação de mineração de dados. Dentre as desvantagens, temos:

- **Custo:** A mineração de dados pode exigir investimentos significativos e custoso em infraestrutura, treinamento, software adequado para garantir que a mineração de dados seja feita de forma apropriada;
- **Informações incorretas ou mal utilizadas:** As técnicas de mineração de dados podem ser utilizadas para coletar informações para objetivos antiéticos. Informações que podem refletir as desigualdades presentes na sociedade;
- **Questões de privacidade:** Pode ser uma desvantagem da mineração de dados as questões de privacidade e segurança dos dados, pois para entender tendências, as organizações tendem a coletar diversos dados dos usuários. É válida a preocupação com a privacidade pessoal dos indivíduos.
- **Dependência de dados de qualidade:** Se os dados forem incompletos ou inconsistentes, os resultados da mineração de dados podem ser comprometidos.

Ao considerar essas desvantagens, é possível adotar medidas para minimizar seus impactos, obtendo melhores resultados em diferentes contextos.

Exemplos de uso interessantes

Uma história de sucesso da aplicação de Mineração de Dados é da empresa Uber. O Uber utiliza a análise e mineração de dados para melhorar a experiência do usuário e tomar decisões estratégicas.

Um caso notável envolve a otimização da precificação dinâmica, conhecida como "surge pricing". O Uber implementou um algoritmo de data mining que analisa uma variedade de fatores, como demanda, tempo estimado de viagem, disponibilidade de motoristas e eventos locais, para determinar os

momentos e locais em que a demanda por corridas é maior.

Com a análise dessas informações, o Uber aplica um preço dinâmico, mais alto durante períodos de maior demanda, incentiva motorista a trabalharem nesses períodos para atender a demanda e equilibrar a oferta e procura. Garantindo também, que os usuários do Uber tenham acesso ao seu serviço quando assim necessitam.

A empresa também utiliza de mineração de dados para analisar dados de viagens, padrões de deslocamento, preferências do usuário e informações demográficas para identificar novas oportunidades de mercado, planejar expansões em novas cidades e desenvolver parcerias estratégicas.

Esse case de sucesso do Uber com a aplicação de Data Mining mostra como a técnica permite que empresas otimizem suas operações e tomem decisões estratégicas para seus serviços.

Já um caso de insucesso, temos da empresa “Target”, que a utilização de mineração de dados causou grande polêmica quando, em 2012, descobriu a gravidez de uma adolescente antes mesmo que ela tivesse contado para sua família. Através de análise de dados da compra de seus produtos, a empresa passou a recomendar alguns produtos relacionados a gravidez com base na compra de produtos que eram frequentemente comprados por mulheres grávidas.

O incidente gerou preocupações da privacidade de dados e até que ponto as empresas podem utilizar dos dados dos consumidores. Isso mostra que uma das desvantagens apresentadas anteriormente, privacidade e segurança e informações incorretas e mal utilizadas. Conclui-se que a utilização da mineração de dados deve também ser cautelosa.

Referências Bibliográficas

BRAZ, Antonio. Mineração de Dados. Recife: UFPE, 20XX. Disponível em: https://www.cin.ufpe.br/~in940/Mineracao_de_Dados-AntonioBraz.pdf. Acesso em: 09 jun. 2023.

TECHTARGET. Data Mining. TechTarget. Disponível em: <https://www.techtarget.com/searchbusinessanalytics/definition/data-mining>. Acesso em: 09 jun. 2023.

TECHOPEDIA. Knowledge Discovery in Databases (KDD). Disponível em: <https://www.techopedia.com/definition/25827/knowledge-discovery-in-databases-kdd>. Acesso em: 09 jun. 2023.

JAVATPOINT. Advantages and Disadvantages of Data Mining. Disponível em: <https://www.javatpoint.com/advantages-and-disadvantages-of-data-mining#:~:text=Reduced%20fraud%20and%20increased%20organizational,bias%20data%20and%20false%20insights>. Acesso em: 10 jun. 2023

GOLDSCHMIDT, Ronaldo; PASSOS, Emmanuel; BEZERRA, Eduardo. Data Mining: Conceitos, Técnicas, Algoritmos, Orientações e Aplicações. eBook Kindle. GEN LTC, 2º edição, 2021.

Base de dados

A base de dados encontrada com nome imdb_js apresenta dados sobre filmes e seus respectivos diretores, autores, personagens e as notas que cada filme recebe. A nota é do site <https://www.imdb.com/> Internet Movie Database. Qualquer usuário logado pode avaliar filmes, séries, documentários, etc.

O objetivo da base é armazenar dados relevantes dos filmes e das avaliações dos usuários, os votos individuais são agregados em uma nota única, nota que é apresentada no site.

Diagrama Entidade-Relacionamento

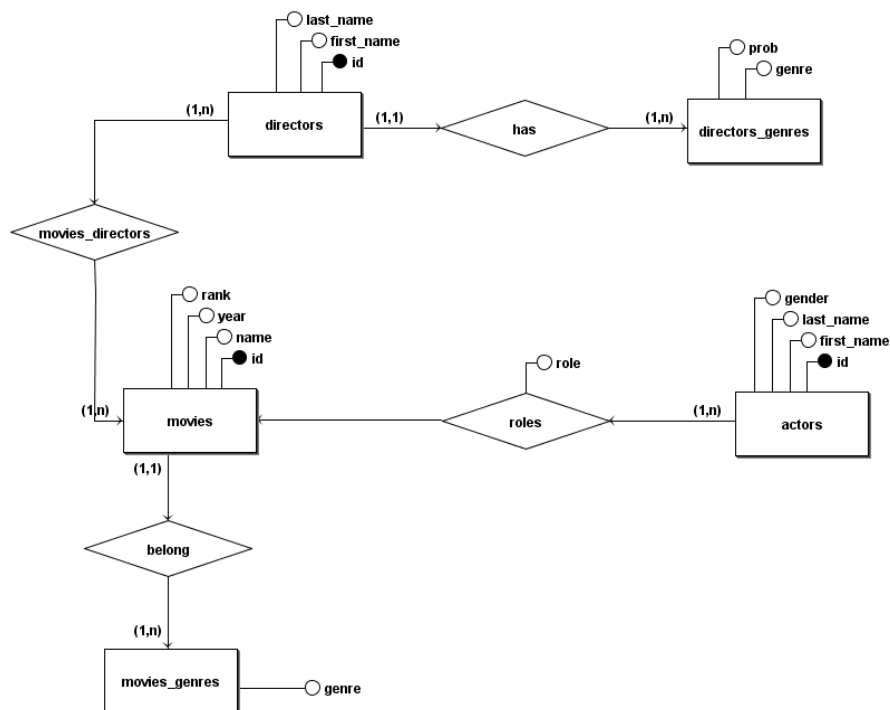
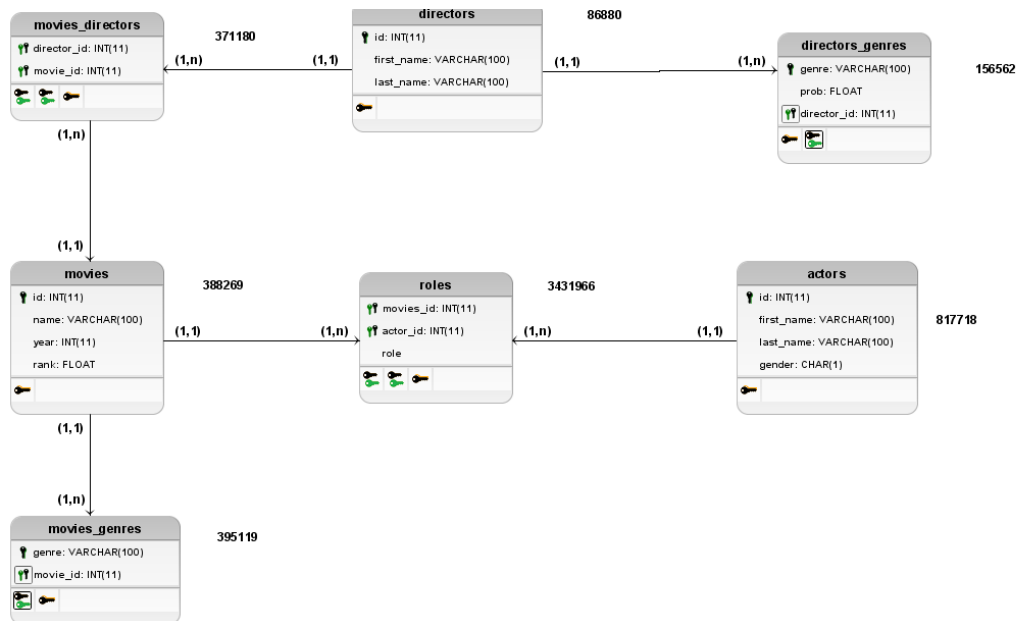


Diagrama Lógico de Dados



A base de dados foi encontrada no seguinte site:

<https://relational.fit.cvut.cz/dataset/IMDb>.

O site apresenta a base de dados e instruções de acesso da base pelo MySQL.