

UNIVERSIDADE DE BRASÍLIA
Faculdade do Gama

Sistemas de Banco de Dados 2

Trabalho Final (TF)
Pesquisa em Data Mining (DM Mineração de Dados)

Nome: Artur Rodrigues Sousa Alves

Matrícula: 211043638

Brasília, DF

2023

a) Definição da Tecnologia Pesquisada

Data Mining, ou Mineração de Dados, representa uma disciplina essencial no campo da análise de dados, destacando-se como uma ferramenta poderosa para extrair conhecimentos valiosos de conjuntos massivos de informações. No cerne dessa prática, encontramos a busca por padrões, relações e tendências que muitas vezes permanecem ocultos em meio ao volume de dados. Seu contexto de utilização abrange diversas áreas, desde a pesquisa científica até aplicações práticas em negócios e tecnologia.

No âmbito mais amplo, Data Mining fundamenta-se em métodos estatísticos, aprendizado de máquina e inteligência artificial, capacitando analistas e sistemas a explorarem dados de maneira mais profunda do que as abordagens tradicionais permitiriam. O principal objetivo é transformar dados brutos em informações estratégicas, fornecendo uma visão mais clara e acionável dos fenômenos subjacentes aos conjuntos de dados.

As definições e conceitos fundamentais envolvem técnicas sofisticadas de análise, incluindo classificação, regressão, clusterização e associação. Cada uma dessas abordagens desempenha um papel crucial na revelação de padrões específicos, permitindo insights valiosos e decisões informadas.

Para exemplificar, no cenário empresarial, a Mineração de Dados é frequentemente aplicada para identificar tendências de mercado, compreender o comportamento do cliente, aprimorar estratégias de marketing e otimizar processos de negócios. Em pesquisa científica, pode ser empregada para descobrir padrões em experimentos, acelerando a obtenção de novos conhecimentos.

Nesta era de explosão de dados, onde a informação é um ativo estratégico, a prática do Data Mining emerge como um componente crítico para transformar grandes volumes de dados em inteligência acionável, contribuindo significativamente para avanços em diversas disciplinas e impulsionando a inovação em diferentes setores.

b) Objetivo(s) principal(is) da Tecnologia Pesquisada

O principal objetivo da tecnologia de Data Mining é extrair conhecimento valioso e previamente desconhecido a partir de grandes conjuntos de dados. Esse processo envolve a identificação de padrões, relações e tendências que não são facilmente discerníveis por meio de métodos de análise convencionais. Alguns dos principais objetivos de Data Mining são:

1. **Descoberta de Padrões e Tendências:** O Data Mining visa identificar padrões significativos nos dados, permitindo a compreensão de tendências ocultas e a revelação de informações valiosas.
2. **Previsão e Análise de Tendências:** Ao analisar padrões passados, o Data Mining é capaz de realizar previsões sobre tendências futuras. Isso é especialmente útil em setores como finanças, marketing e saúde.
3. **Segmentação de Dados e Clientes:** A tecnologia visa dividir grandes conjuntos de dados em segmentos mais gerenciáveis, permitindo uma compreensão mais aprofundada de diferentes grupos de clientes, padrões de comportamento e preferências.
4. **Classificação e Categorização:** Classificar dados em categorias específicas é um objetivo chave do Data Mining. Isso pode ser aplicado em diversos contextos, como classificação de e-mails como spam ou não spam, diagnóstico médico e muito mais.
5. **Otimização de Processos de Negócios:** Ao analisar dados operacionais, o Data Mining procura otimizar processos de negócios, identificando áreas de melhoria, reduzindo custos e aumentando a eficiência.
6. **Identificação de Anomalias:** A detecção de anomalias em conjuntos de dados é crucial para diversos setores, onde a identificação de comportamentos fora do padrão pode indicar ameaças ou problemas.

7. Apoio à Tomada de Decisões: Fornecer informações acionáveis para apoiar a tomada de decisões é um dos objetivos fundamentais do Data Mining. Essas informações podem guiar estratégias de negócios, políticas públicas, tratamentos médicos e muito mais.

8. Personalização e Recomendação: Em setores como comércio eletrônico e entretenimento, o Data Mining é usado para personalizar experiências do usuário e fazer recomendações relevantes, aumentando a satisfação do cliente.

Em suma, o objetivo primordial do Data Mining é transformar dados brutos em conhecimento prático, proporcionando uma visão mais profunda e estratégica dos fenômenos subjacentes. Essa tecnologia desempenha um papel crucial na era da informação, capacitando organizações e profissionais a tomar decisões mais informadas e eficazes.

c) Vantagens da Tecnologia Pesquisada

A tecnologia de Data Mining oferece uma série de vantagens significativas, proporcionando insights valiosos a partir de grandes conjuntos de dados. Aqui estão algumas das principais vantagens dessa prática:

1. Descoberta de Padrões Complexos: Sendo tanto objetivo quanto vantagem o Data Mining permite a identificação de padrões complexos e não triviais nos dados, que podem não ser aparentes por meio de métodos de análise convencionais.

2. Tomada de Decisões Informada: As análises geradas pelo Data Mining fornecem informações acionáveis que sustentam a tomada de decisões informada em diversas áreas, incluindo negócios, saúde, finanças e governo.

3. Previsão de Tendências: Sendo tanto objetivo quanto vantagem ao analisar dados históricos, o Data Mining pode prever tendências futuras e comportamentos, permitindo que organizações se antecipem a mudanças no mercado.

4. Identificação de Oportunidades de Negócios: Ao compreender padrões de compra, preferências do cliente e comportamento do mercado, as empresas podem identificar oportunidades de negócios lucrativas.

5. Segmentação de Mercado: O Data Mining facilita a segmentação eficaz de mercados, permitindo que as empresas personalizem estratégias de marketing e atendam de forma mais eficaz às necessidades específicas de diferentes grupos de clientes.

6. Detecção de Anomalias e Fraudes: A tecnologia é eficaz na detecção de atividades anômalas nos dados, sendo fundamental para a prevenção de fraudes em setores como finanças e seguros.

7. Otimização de Processos Operacionais: Ao analisar dados operacionais, o Data Mining pode identificar ineficiências e otimizar processos, resultando em maior eficiência operacional e redução de custos.

8. Personalização de Experiências do Cliente: Empresas podem usar Data Mining para personalizar experiências do cliente, fornecendo recomendações precisas e relevantes com base no histórico de comportamento do usuário.

9. Aprimoramento da Qualidade dos Produtos e Serviços: Ao entender as preferências dos clientes e feedbacks, as organizações podem aprimorar continuamente a qualidade de produtos e serviços.

10. Aumento da Competitividade: Empresas que aproveitam o Data Mining ganham uma vantagem competitiva, pois estão melhor posicionadas para tomar decisões estratégicas e adaptar-se às mudanças do mercado.

Em resumo, o Data Mining oferece vantagens substanciais, contribuindo para a inovação, eficiência e sucesso organizacional em uma variedade de setores. Essas vantagens tornam essa tecnologia uma ferramenta valiosa para aproveitar o poder dos dados na era da informação.

d) Desvantagens da Tecnologia Pesquisada

Embora o Data Mining ofereça uma série de benefícios, é importante reconhecer suas desvantagens e desafios associados. Aqui estão algumas das principais desvantagens da tecnologia de Data Mining:

1. Privacidade e Ética: O uso extensivo de dados para análise pode levantar preocupações com a privacidade. A coleta e análise de informações pessoais podem resultar em questões éticas, especialmente se não forem implementadas práticas robustas de proteção de dados.
2. Dependência de Dados de Qualidade: A eficácia do Data Mining depende da qualidade dos dados utilizados. Dados imprecisos, desatualizados ou incompletos podem levar a conclusões errôneas e impactar negativamente as decisões organizacionais.
3. Complexidade e Interpretação: Os algoritmos de Data Mining muitas vezes geram modelos complexos e difíceis de interpretar. Isso pode ser um desafio, especialmente quando os resultados precisam ser explicados a não especialistas.
4. Custo de Implementação: A implementação efetiva de soluções de Data Mining pode envolver custos significativos. Isso inclui a aquisição de software especializado, treinamento de equipe e manutenção contínua.
5. Risco de Viés nos Dados: Se os dados utilizados para treinar modelos de Data Mining forem enviesados, os resultados também serão enviesados. Isso

pode levar a decisões discriminatórias ou não equitativas.

6. Segurança da Informação: O aumento do acesso e manipulação de grandes conjuntos de dados pode aumentar os riscos de segurança. A proteção adequada contra ameaças cibernéticas é crucial para garantir a integridade dos dados.

7. Limitações em Dados Não Estruturados: Embora seja eficaz em dados estruturados, o Data Mining pode ter dificuldades em lidar com dados não estruturados, como texto livre e multimídia.

8. Necessidade de Especialistas: A implementação bem-sucedida de Data Mining muitas vezes requer especialistas em estatísticas, aprendizado de máquina e análise de dados. A falta de profissionais qualificados pode ser um obstáculo.

9. Risco de Overfitting: Modelos de Data Mining podem se ajustar demais aos dados de treinamento, resultando em overfitting. Isso significa que o modelo pode ser muito específico para os dados de treinamento e não generalizar bem para novos dados.

10. Aceitação Organizacional: Algumas organizações podem enfrentar resistência interna à implementação de soluções de Data Mining, especialmente se houver preocupações sobre a substituição de processos existentes ou o temor da automação de decisões.

11. Tempo e Recursos: A análise de grandes conjuntos de dados pode ser intensiva em termos de tempo e recursos computacionais, especialmente ao lidar com volumes massivos de informações.

Embora o Data Mining ofereça inúmeras vantagens, é fundamental considerar essas desvantagens ao implementar essa tecnologia, garantindo uma abordagem equilibrada e ética na utilização de dados para tomada de decisões.

e) Exemplo(s) de uso interessante(s) em empresas, projetos ou instituições dessa tecnologia de Banco de Dados pesquisada

1. Netflix: Personalização de Recomendações:

A Netflix é conhecida por utilizar extensivamente técnicas de Data Mining para aprimorar a experiência do usuário. O algoritmo de recomendação da Netflix analisa os hábitos de visualização, preferências e avaliações dos usuários para sugerir filmes e séries personalizadas. Esse uso eficaz de Data Mining contribui significativamente para a retenção de assinantes e a satisfação do cliente. Esse uso é descrito no objeto **8** do tópico **c)** de pesquisa.

2. Amazon: Sistemas de Recomendação:

A Amazon é pioneira no uso de Data Mining para impulsionar seus sistemas de recomendação. Ao analisar o histórico de compras, padrões de navegação e avaliações dos clientes, a Amazon fornece recomendações altamente personalizadas, aumentando as chances de venda cruzada e melhorando a experiência do cliente. Esse uso é descrito nos objetos **8** e **9** do tópico **c)** de pesquisa.

3. Banco Santander: Detecção de Fraudes:

Instituições financeiras, como o Banco Santander, empregam Data Mining para detecção proativa de fraudes. Analisando padrões de transações, comportamentos atípicos e anomalias, o banco pode identificar potenciais atividades fraudulentas, protegendo os clientes e a integridade financeira. Esse uso é descrito no objeto **6** do tópico **c)** de pesquisa.

Esses exemplos destacam como o Data Mining se tornou uma ferramenta essencial para empresas em diversas indústrias, demonstrando seu potencial em impulsionar a personalização, eficiência operacional e tomada de decisões informada. Essas aplicações bem-sucedidas reforçam a importância do Data Mining no cenário empresarial moderno.

f) Bibliografias Pesquisadas

Data mining. TechTarget, 2023. Disponível em: <<https://www.techtarget.com/searchbusinessanalytics/definition/data-mining>>. Acesso em: 14 de novembro de 2023.

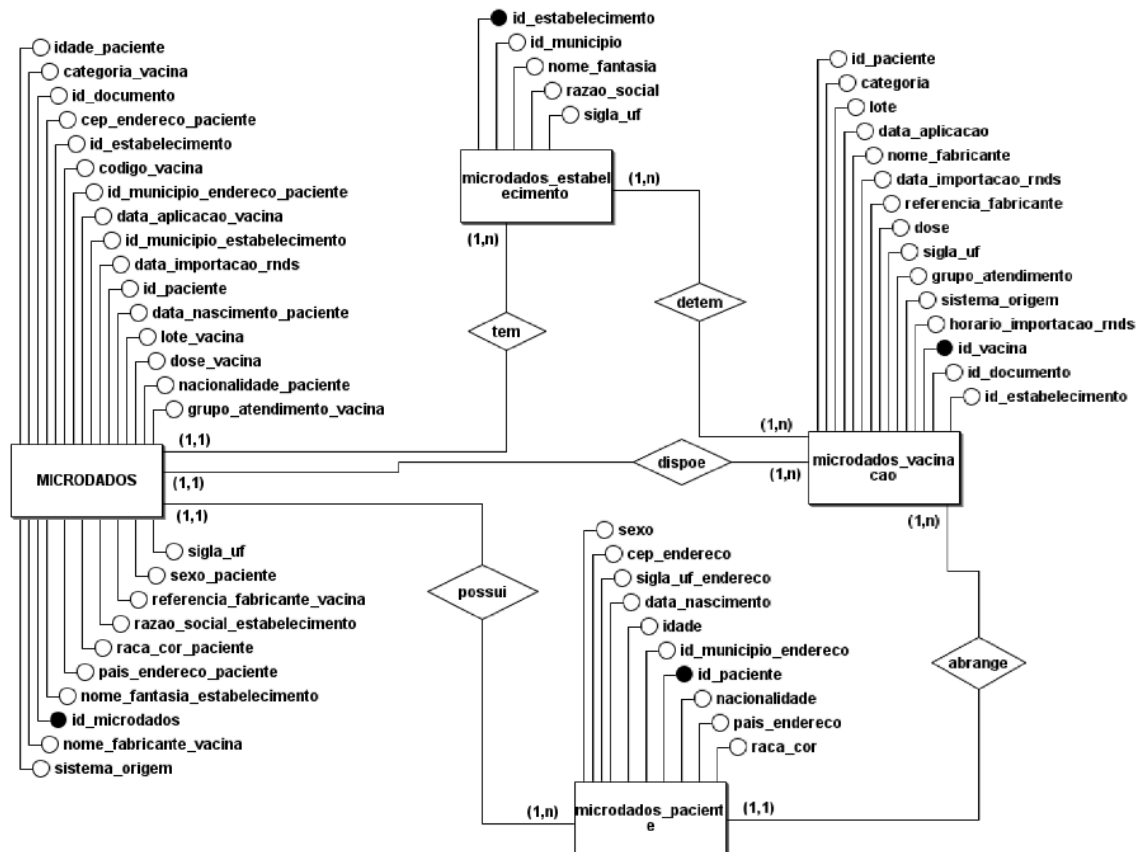
AMARAL, Fernando. Mineração de dados: conceitos e aplicações. In: Aprenda mineração de dados: teoria e prática. 1. ed. São Paulo: Disal, 2023.

FUNDAÇÃO INSTITUTO DE ADMINISTRAÇÃO (FIA). Data Mining: Descubra como essa técnica pode transformar seus negócios. Disponível em: <https://fia.com.br/blog/data-mining/>. Acesso em: 14 de novembro de 2023.

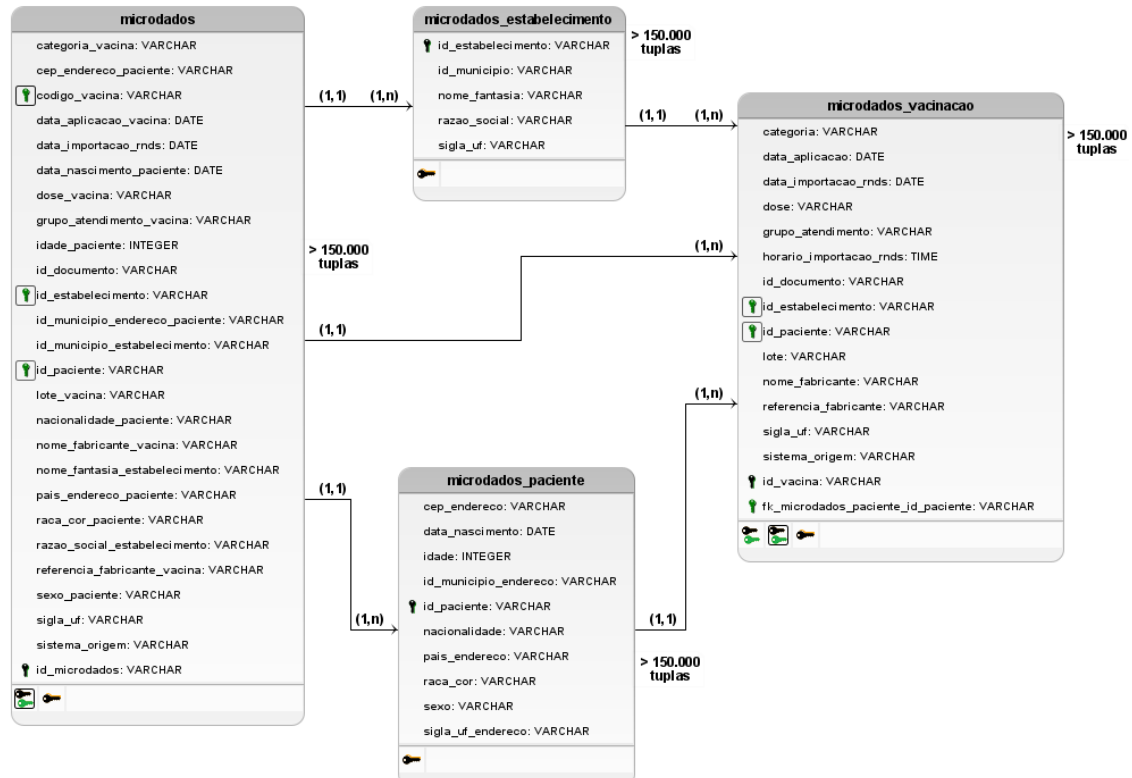
g) Base de Dados (documentação)

Vemos, nessa base de dados, a disponibilização de dados referentes à Campanha Nacional de Vacinação contra Covid-19 pelo Ministério da Saúde, através do Sistema de Informação do Programa Nacional de Imunizações (SI-PNI). O objetivo da base é oferecer registros de vacinação anonimizados, provenientes da Rede Nacional de Dados em Saúde (RNDS), desde o início da campanha. Esses dados incluem informações sobre o número de doses aplicadas, desagregadas por unidade federativa, município, período, gênero, faixa etária e tipo de vacina. Também estão disponíveis percentuais de vacinados, coberturas vacinais e mapas de distribuição das coberturas vacinais. O contexto envolve a disponibilização de dados cruciais para a monitorização e análise da vacinação contra a Covid-19 no Brasil, permitindo que instituições e o público em geral acessem e utilizem esses dados para tomadas de decisão e estudos epidemiológicos.

DE-R (Diagrama Entidade-Relacionamento)



DLD (Diagrama Lógico)



Dicionário de Dados

Entidade: MICRODADOS				
Descrição: Microdados paralelos ao dado original, incluindo pacientes, vacinação e estabelecimentos.				
Atributo	Propriedades do atributo	Tipo de dado	Tamanho	Descrição
categoria_vacina	obrigatório	varchar	-	Categoria
cep_endereco_paciente	obrigatório	varchar	-	CEP do Endereço (5 dígitos para anonimizado e 7 dígitos para identificado)
codigo_vacina	Chave estrangeira, obrigatório	varchar	-	Código da vacina
data_aplicacao_vacina	obrigatório	date	-	Data de aplicação
data_importacao_rnds	obrigatório	date	-	Data de importação
data_nascimento_paciente	obrigatório	date	-	Data de nascimento
dose_vacina	obrigatório	varchar	-	Dose
grupo_atendimento_vacina	obrigatório	varchar	-	Grupo de atendimento
idade_paciente	obrigatório	integer	-	Idade (ano)
id_documento	obrigatório	varchar	-	ID Documento
Id_estabelecimento	Chave estrangeira, obrigatório	varchar	-	ID Estabelecimento - CNES
id_municipio_endereco_paciente	obrigatório	varchar	-	ID Município do Endereço - IBGE 7 Dígitos
id_municipio_estabelecimento	obrigatório	varchar	-	ID Município - IBGE 7 Dígitos
id_paciente	Chave estrangeira, obrigatório	varchar	-	ID Paciente
lote_vacina	obrigatório	varchar	-	Número do Lote
nacionalidade_paciente	obrigatório	varchar	-	Nacionalidade
nome_fabricante_vacina	obrigatório	varchar	-	Nome do fabricante/fornecedor
nome_fantasia_estabelecimento	obrigatório	varchar	-	Nome Fantasia

nto				
pais_endereco_paciente	obrigatório	varchar	-	País do Endereço
raca_cor_paciente	obrigatório	varchar	-	Raça ou cor
razao_social_estabelecimento	obrigatório	varchar	-	Razão Social
referencia_fabricante_vacina	obrigatório	varchar	-	CNPJ do fabricante/fornecedor
sexo_paciente	obrigatório	varchar	-	Sexo
sigla_uf	obrigatório	varchar	-	Sigla da Unidade da Federação
sistema_origem	obrigatório	varchar	-	Sistema de origem
Id_microdados	Chave primária obrigatória	varchar	-	ID microdados

Entidade: microdados_estabelecimento
Descrição: Microdados normalizados de estabelecimentos

Atributo	Propriedades do atributo	Tipo de dado	Tamanho	Descrição
id_estabelecimento	chave primária obrigatório	varchar	-	ID Estabelecimento - CNES
id_municipio	obrigatório	varchar	-	ID Município - IBGE 7 Dígitos
nome_fantasia	obrigatório	varchar		Nome Fantasia
razao_social	obrigatório	varchar	-	Razão Social
sigla_uf	obrigatório	varchar	-	Sigla da Unidade da Federação

Entidade: microdados_paciente
Descrição: Microdados normalizados de pacientes

Atributo	Propriedades do atributo	Tipo de dado	Tamanho	Descrição
cep_endereco	chave primária obrigatório	varchar	-	CEP do Endereço (5 dígitos para anonimizado e 7 dígitos para identificado)
data_nascimento	obrigatório	varchar	-	Data de nascimento
idade	obrigatório	integer		Idade (ano)

id_municipio_endereco	obrigatório	varchar	-	ID Município do Endereço - IBGE 7 Dígitos
id_paciente	obrigatório	varchar	-	ID Paciente
nacionalidade		varchar	-	Nacionalidade
pais_endereco		varchar	-	País do Endereço
raca_cor		varchar	-	Raça ou cor
sexo		varchar	-	Sexo
sigla_uf_endereco		varchar	-	Sigla da Unidade da Federação do Endereço

Entidade: microdados_vacinacao				
Descrição: Microdados normalizados de vacinação				
Atributo	Propriedades do atributo	Tipo de dado	Tamanho	Descrição
categoria	obrigatório	varchar	-	Categoria
data_aplicacao	obrigatório	varchar	-	Data de aplicação
data_importacao_rnds	obrigatório	integer		Data de importação
dose	obrigatório	varchar	-	Dose
grupo_atendimento	obrigatório	varchar	-	Grupo de atendimento
horario_importacao_rnds		varchar	-	Horário de importação
id_documento		varchar	-	ID Documento
id_estabelecimento	Chave estrangeira obrigatória	varchar	-	ID Estabelecimento - CNES
id_paciente	Chave estrangeira obrigatória	varchar	-	ID Paciente
lote	obrigatório	varchar	-	Número do Lote
nome_fabricante	obrigatório	varchar	-	Nome do fabricante/fornecedor
referencia_fabricante	obrigatório	varchar	-	CNPJ do fabricante/fornecedor
sigla_uf	obrigatório	varchar	-	Sigla da Unidade da Federação

sistema_origem	obrigatório	varchar	-	Sistema de origem
id_vacina	Chave primária obrigatória	varchar	-	Vacina

Relacionamento: POSSUI				
Descrição: Define relacionamento que corresponde ao relacionamento entre as entidades MICRODADOS e MICRODADOS_PACIENTE; cardinalidade 1:n				
Atributo	Propriedades do atributo	Tipo de dado	Tamanho	Descrição
id_microdados	chave estrangeira chave candidata composta obrigatório	varchar	-	ID Microdados
id_paciente	chave estrangeira chave candidata composta obrigatório	varchar	-	ID Paciente

Relacionamento: DISPOE				
Descrição: Define relacionamento que corresponde ao relacionamento entre as entidades MICRODADOS e MICRODADOS_VACINACAO; cardinalidade 1:n				
Atributo	Propriedades do atributo	Tipo de dado	Tamanho	Descrição
id_microdados	chave estrangeira chave candidata composta obrigatório	varchar	-	ID Microdados
id_vacina	chave estrangeira chave candidata composta obrigatório	varchar	-	ID Vacina

Relacionamento: TEM				
Descrição: Define relacionamento que corresponde ao relacionamento entre as entidades MICRODADOS e MICRODADOS_ESTABELECIMENTO; cardinalidade 1:n				
Atributo	Propriedades do atributo	Tipo de dado	Tamanho	Descrição
id_microdados	chave estrangeira chave candidata composta obrigatório	varchar	-	ID Microdados
id_estabelecimento	chave estrangeira chave candidata composta obrigatório	varchar	-	ID Estabelecimento

Relacionamento: DETEM				
Descrição: Define relacionamento que corresponde ao relacionamento entre as entidades MICRODADOS_VACINACAO e MICRODADOS_ESTABELECIMENTO; cardinalidade 1:n				
Atributo	Propriedades do atributo	Tipo de dado	Tamanho	Descrição
id_vacina	chave estrangeira chave candidata composta obrigatório	varchar	-	ID Vacina
id_estabelecimento	chave estrangeira chave candidata composta obrigatório	varchar	-	ID Estabelecimento

Relacionamento: ABRANGE				
Descrição: Define relacionamento que corresponde ao relacionamento entre as entidades MICRODADOS_VACINACAO e MICRODADOS_PACIENTE; cardinalidade 1:n				
Atributo	Propriedades do atributo	Tipo de dado	Tamanho	Descrição
id_vacina	chave estrangeira chave candidata composta obrigatório	varchar	-	ID Vacina

Endereço virtual específico da base de dados:

<https://opendatasus.saude.gov.br/dataset/covid-19-vacinacao>

<https://basedosdados.org/dataset/5020ca16-6df4-43bd-8717->

[eab37ef84c30?table=364c115e-b3b5-471d-be44-1c15cfe97f40](https://basedosdados.org/dataset/5020ca16-6df4-43bd-8717-eab37ef84c30?table=364c115e-b3b5-471d-be44-1c15cfe97f40)