

Desde a década de 1970 já se falava em sistemas de apoio a tomadas de decisão aplicadas. Mas foi com o advento da internet, nas décadas de 1980 e 1990, que a ideia começou a ganhar força, dado o grande volume de dados gratuitos e de fácil acesso à rede. Demonstrado no artigo de **Jing He**, “**O processo data mining desenvolveu-se devido ao imenso volume de dados que devem ser tratados mais facilmente em áreas como: negócios, indústria médica, astronomia, genética ou área bancária.**” {2}.

Big Data é um termo utilizado para definir grandes e complexos conjuntos de dados. Sendo atualmente muito usado nas áreas de negócios e pesquisa, como diz o autor **Pedro Neves** “**Big Data é um tema atual entre pesquisa e negócios áreas cujos desenvolvimentos foram encorajados pelas altas demandas de armazenamento e capacidade de processamento necessário para lidar grandes conjuntos de dados.**” {4}.

No artigo {5} de **Nawsher Khan**, ele resume Big data em um modo geral nas suas características, problemas e desafios, pode ser resumida em 10Vs, esses 10 Vs seria:

- Volume
- Valor
- Velocidade
- Veracidade
- Variabilidade
- Volatilidade
- Viabilidade
- Validade
- Variedade

Os 3Vs mais importantes seriam:

- Volume pela grande quantidade de dados a ser processado
- Velocidade para analisar os dados em tempo satisfatório
- Variedade devido aos diferentes tipos de dados utilizados.

Mineração de dados (em inglês, data mining), é um método de análise profunda que abrange uma enorme quantidade de dados de informações. Por isso, o sistema tem que contar com uma capacidade robusta para coletar e analisar dados com objetivo de encontrar anomalias, padrões e correlações em grandes conjuntos de dados para prever resultados. O sistema responsável por organizar e estudar as informações colhidas precisa apresentar algumas características fundamentais, como contar com diferentes formas de classificação e preparação de dados.

Com o passar dos anos o numero de dados produzidos está aumentando consideravelmente. Tanto os dados Estruturados aqueles que são armazenados em bancos de dados tradicionais, os Semiestruturados, que seguem diversos padrões, de formas heterógenas e o Não Estruturados, compõem 90% do universo digital, como vídeos, texto, áudio e imagens no artigo de K Sriraghav demonstra que “**existe a necessidade de uma ferramenta de ponta a ponta que possa estruturar, resumir e visualizar dados não estruturados.**” {3} Com isso, a mineração de dados separe todos os ruídos caóticos e repetitivos em seus dados, entenda o que é relevante para, então, fazer um bom uso dessa informação para avaliar os resultados possíveis e acelere o ritmo de tomadas de decisões bem-informadas.

Esse processo envolve técnicas e métodos de diversas áreas como: inteligência artificial, aprendizado de máquina, estatísticas e sistemas de banco de dados. O objetivo aqui é extrair dados de fontes de dados e transformá-los em informação útil, gerando valor para o negócio.

Varejistas, empresas, bancos, fabricantes, operadoras de telecomunicações, seguradoras, estão usando a mineração de dados para descobrir relações entre tudo desde preços, promoções e demografias até como a economia, o risco, a concorrência e as mídias sociais estão afetando seus modelos de negócio, receitas, operações e relacionamentos com os clientes. Exemplo na prática, na atualidade: A receita federal utiliza muita a mineração de dados no caso da Lava Jato em confusão com a polícia federal. O processo envolvido são algoritmos, tecnologia da informação, programas.

Podemos comprovar isso com o artigo de **SHIVAM AGARWAL** que diz “O principal objetivo do processo de mineração de dados é extrair informações úteis do dossiê de dados e moldá-lo em uma estrutura compreensível para uso futuro. Existem diferentes processos e técnicas usadas para realizar a mineração de dados com sucesso.”. {1}

Mineração de dados não é apenas um processo computacional ou automático, mas sim um trabalho cooperativo entre pessoas e máquinas. Pessoas conseguem descrever, modelar e desenhar bancos de dados, os problemas a serem solucionados e os objetivos a serem alcançados. Máquinas processam dados e pessoas buscam soluções para os problemas usando a informação gerada.

Os maiores benefícios da mineração de dados é a criação de inteligência de negócios sobre determinado assunto, também referenciado como: KDD – Knowledge Discovery in Databases – ou Descoberta de Conhecimento sob Base de Dados.

Existem diversas situações onde podemos aplicar a Mineração de dados para resolver problemas importantes.

- Gestão e vendas
- Tecnologia
- Administração e marketing
- Educação
- Saúde

Mineração de dados e Big Data é um dos assuntos mais falados atualmente. Cada vez mais, profissionais e pesquisadores buscam aprimorar seus conhecimentos no assunto a fim de dominarem essa área.

Aprofundado na área da saúde, a mineração de dados médicos é um processo desafiador. A falta de grandes bases de dados e a complexidade dos dados são alguns dos desafios da complexidade das bases médicas, que podem ser compostas por várias fontes como dados de exames laboratoriais, entrevistas e imagens, no artigo de **Riccardo Bellazzi** comprova que a mineração de dados é de extrema importância na área da saúde “**mineração de dados oferece soluções metodológicas e técnicas para lidar com a análise de dados médicos e a construção de modelos de previsão. Uma grande variedade desses métodos requer diretrizes gerais e simples que podem ajudar os profissionais na seleção apropriada de ferramentas de mineração de dados, construção e validação de modelos preditivos, juntamente com a disseminação de modelos preditivos em ambientes clínicos.**”. {6}

Muitas vezes, os pacientes não realizam os procedimentos e exames requisitados e deixam lacunas nos prontuários. As bases de dados de doenças, principalmente, provêm de diversas fontes

diferentes, como entrevista com o paciente, testes laboratoriais, resultados de equipamentos e exames diretos. Isso tende a produzir bases de dados altamente variadas e difíceis de serem analisadas, demandando o uso de diferentes técnicas e ferramentas para serem exploradas de maneira eficiente.

Apesar de grandes avanços na área de informática e gerenciamento de dados da saúde no que se refere a estatísticas, gerenciamento de grandes hospitais e estudos de larga escala, há uma escassez de ferramentas analíticas para se extrair conhecimento desses dados. No entanto, a mineração de dados vem ganhando espaço na área da saúde, **Mohamed Elhoseny** comprova isso em seu artigo **“Na última década, tem havido um crescente interesse em pesquisa de big data, especialmente para aplicações de serviços de saúde. A adoção do paradigma da computação em nuvem e da Internet das Coisas (IoT) na área da saúde pode trazer várias oportunidades à TI médica, e os especialistas acreditam que ela pode melhorar significativamente os serviços de saúde e contribuir para sua inovação contínua e sistemática em um ambiente de big data como aplicativos do Industry 4.0. No entanto, os recursos necessários para gerenciar esses dados em um ambiente de nuvem-IoT ainda são um grande desafio”{7}.**

Uma das grandes vantagens de usar a mineração em dados é que ela pode ser usada pelas operadoras de planos de saúde para detectar fraudes e abusos, **Melih Kirlidog** explica como isso pode acontecer em seu artigo **“A fraude pode ser vista em todos os tipos de seguro, incluindo seguro de saúde. A fraude no seguro de saúde é feita por engano intencional ou deturpação de ganhar algum benefício miserável na forma de gastos com saúde. Ferramentas e técnicas de mineração de dados podem ser usado para detectar fraudes em grandes conjuntos de dados de pedidos de seguro. Com base em alguns casos conhecidos ou suspeitos de serem fraudulentos, a técnica de detecção de anomalias calcula a probabilidade ou probabilidade de cada registro ser fraudulento, analisando o passado reclamações de seguro. Os analistas podem, então, ter uma investigação mais detalhada dos casos que foram marcados pelo software de mineração de dados.”{8}.**

**Contudo** a principal vantagem é ajudar as organizações de saúde a tomar decisões de gerenciamento e de relação com clientes. Médicos podem identificar tratamentos eficazes e boas práticas clínicas, comprovado no artigo de **Nada Lavrac** **“A importância da interpretabilidade de os resultados da análise de dados são discutidos e ilustrados em aplicações médicas selecionadas.” {9}.**

Pode ainda ajudar pesquisadores a identificar sintomas e características da doença para diagnóstico e tratamento. Com isso, novas descobertas biológicas estão acontecendo, como no artigo de **Peng Jiang & X Shirley Liu** que demonstra, **“Nos últimos anos, o rápido crescimento de dados biológicos em larga escala, mas a efetiva mineração e modelagem de “big data” para novas descobertas biológicas continua sendo um desafio significativo. Um novo estudo reanalisa os perfis de expressão do Gene Expression Omnibus para fazer novas descobertas sobre genes envolvidos no reparo de danos no DNA e na instabilidade do genoma no câncer.”{11}.**

Médicos tomam decisões diagnósticas e recomendam de tratamentos baseados no histórico do paciente, exames clínicos, e laboratoriais. Aplicar técnicas de mineração de dados nas bases médicas pode fornecer aos médicos ferramentas analíticas e preditivas que vão além do que é visto na superfície dos dados

A mineração de dados vem ajudando na prevenção e no tratamento de diversas doenças como no caso da diabetes, hoje, no Brasil, há mais de 13 milhões de pessoas vivendo com diabetes, o que representa 6,9% da população. E esse número está crescendo. Em alguns casos, o diagnóstico demora, favorecendo o aparecimento de complicações.

No artigo de **Ioannis Kavakioti** demonstra que “Os avanços notáveis em biotecnologia e ciências da saúde levaram a uma produção significativa de dados, como dados genéticos de alto rendimento e informações clínicas, gerados a partir de grandes registros eletrônicos de saúde. Para este fim, a aplicação de métodos de aprendizado de máquina e mineração de dados em biociências é hoje, mais do que nunca, vital e indispensável nos esforços para transformar de maneira inteligente todas as informações disponíveis em conhecimento valioso. Diabetes mellitus (DM) é definido como um grupo de distúrbios metabólicos que exercem pressão significativa sobre a saúde humana em todo o mundo. Extensa pesquisa em todos os aspectos do diabetes (diagnóstico, etio patofisiologia, terapia, etc.) levou à geração de enormes quantidades de dados. O objetivo do presente estudo é realizar uma revisão sistemática das aplicações de aprendizado de máquina, técnicas de mineração de dados e ferramentas no campo da pesquisa sobre diabetes a respeito de:

- a) Predição e Diagnóstico,
- b) Complicações Diabéticas,
- c) Antecedentes Genéticos e Meio Ambientes,
- e) Cuidados de Saúde e Gestão,

Com a primeira categoria a ser a mais popular. Uma ampla gama de algoritmos de aprendizado de máquina foi empregada. Em geral, 85% dos utilizados foram caracterizados por abordagens de aprendizagem supervisionada e 15% por métodos não supervisionados e, mais especificamente, por regras de associação. As máquinas de vetores de suporte (SVM) surgem como o algoritmo mais bem-sucedido e amplamente utilizado. Em relação ao tipo de dados, os conjuntos de dados clínicos foram usados principalmente. As aplicações de título nos artigos selecionados projetam a utilidade de extrair conhecimento valioso, levando a novas hipóteses visando uma compreensão mais profunda e uma investigação mais aprofundada em DM.”{10}.

A mineração de dados médicos é desafiadora e distinta de outras áreas. Para doenças raras e negligenciadas as dificuldades são agravadas. Apesar dos desafios, a mineração de dados médicos pode ser a mais recompensadora. Achar uma solução para uma pergunta médica relevante pode significar a melhoria da saúde e qualidade de vida de vários pacientes.

Portanto, A mineração de dados ganhou proeminência nas últimas duas décadas como uma disciplina própria que oferece benefícios em relação a muitos domínios, tanto comerciais quanto acadêmicos. Um dos objetivos para o futuro em mineração de dados é a capacidade de poder trabalhar com um numero cada vez maior de dados, comprovado no artigo de **FRANS COENEN** “Outro driver para pesquisa em mineração de dados é o tamanho cada vez maior dos dados com os quais desejamos trabalhar. Estamos, portanto, também interessados em técnicas para meus conjuntos de dados cada vez maiores (e uma variedade cada vez maior de dados).” {12}.

\item \$\cdots\$ (usando para numerar as perguntas)