

NOME DO DOCUMENTO

**Big\_Data (1).pdf**

NÚMERO DE PALAVRAS

**4583 Words**

NÚMERO DE CARACTERES

**25483 Characters**

NÚMERO DE PÁGINAS

**9 Pages**

TAMANHO DO ARQUIVO

**429.8KB**

DATA DE ENVIO

**Jan 15, 2024 6:38 PM GMT**

DATA DO RELATÓRIO

**Jan 15, 2024 6:39 PM GMT****● 58% geral de similaridade**

O total combinado de todas as correspondências, incluindo fontes sobrepostas, para cada banco de dados

- 55% Banco de dados da Internet
- Banco de dados do Crossref
- 15% Banco de dados de trabalhos enviados
- 6% Banco de dados de publicações
- Banco de dados de conteúdo publicado no Crossref

# Big Data

INÊS SANTOS, Universidade da Beira Interior, Portugal

PATRÍCIA MARCOS, Universidade da Beira Interior, Portugal

O presente trabalho procura proporcionar uma visão abrangente e aprofundada no campo do Big Data, examinando as suas tecnologias fundamentais, desafios inerentes e aplicações práticas. São abordados os sistemas de gerenciamento de banco de dados e ferramentas essenciais para o processamento eficiente de grandes volumes de dados. Explora-se também as interseções entre Big Data e outras áreas-chave como Machine Learning, Processamento de Linguagem Natural e Blockchain, destacando como essas disciplinas convergem para impulsionar a análise de dados em larga escala. A segurança e a governança de dados emergem como áreas críticas, com uma análise dos desafios e das melhores práticas associadas à proteção de informações sensíveis. Além disso, são examinados casos de uso em setores diversos, demonstrando como o Big Data está a transformar radicalmente a tomada de decisões e a inovação em várias indústrias. Por fim, o estudo conclui com uma reflexão sobre a importância da ética no uso de Big Data, considerando questões de privacidade e responsabilidade.

## ACM Reference Formath:

Inês Santos and Patrícia Marcos. 2023. Big Data. 1, 1 (December 2023), 9 pages.

## 1 INTRODUÇÃO

No contexto da era digital, a expressão *Big Data* tornou-se um elemento central nas discussões sobre a transformação e a evolução tecnológica. Este termo refere-se a conjuntos vastos e complexos de dados que ultrapassam a capacidade das ferramentas tradicionais de processamento para serem eficazmente captados e processados. Neste cenário, empresas, organizações e setores diversos procuram compreender e tirar proveito dessas vastas quantidades de informações para obter *insights* valiosos, otimizar processos e impulsionar a inovação. Neste trabalho, exploraremos os fundamentos do *Big Data*, examinando as suas origens, dimensões e implicações, enquanto destacamos o papel crucial que desempenha na moldagem do panorama empresarial e tecnológico contemporâneo.

## 2 QUE SÃO OS 5 V'S DO BIG DATA

Os 5 V's do *Big Data* referem-se a cinco características-chave que ajudam a descrever a natureza e os desafios associados ao processamento e gestão de grandes conjuntos de dados. Quanto mais dados são gerados, maior é o esforço para extrair informações. Os centros de dados tiveram que aprender a lidar com o crescimento exponencial de dados gerados. O conceito de big data foi definido inicialmente por 3 V's, mas a literatura mostrou que o seu conceito pode ser expandido para 5 V's, representados pelos seguintes conceitos:

Authors' addresses: Inês Santos, ines.matos.santos@ubi.pt, Universidade da Beira Interior, Covilhã, Portugal; Patrícia Marcos, patricia.marcos@ubi.pt, Universidade da Beira Interior, Covilhã, Portugal.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from [permissions@acm.org](mailto:permissions@acm.org).

© 2023 Association for Computing Machinery.

Manuscript submitted to ACM

Manuscript submitted to ACM

## 2.1 <sup>4</sup> Volume

Relacionado à grande quantidade de dados gerados, devido a extensa variedade de fontes de dados, onde as empresas os guardam. Antes, <sup>4</sup> armazenar essa grande quantidade de dados e informações era um problema, principalmente quando se tratava de processamento, armazenamento e organização, devido ao alto custo dos equipamentos e à pouca oferta. <sup>10</sup> Com a expansão diária dos servidores, ficou muito mais fácil armazenar esses dados, principalmente com a compressão de arquivos e a possibilidade de disponibilizá-los na nuvem. Já a análise torna-se cada vez mais simples, principalmente com o uso de ferramentas próprias para o *Big Data*.

## 2.2 Velocidade

<sup>11</sup> Devido ao grande volume e variedade de dados, todo o processamento deve ser ágil para gerar as informações necessárias. Os dados são cada vez mais procurados em tempo real, com o menor delay possível. <sup>10</sup> Cada vez mais fontes fornecem dados, o que torna necessário o gestão dessas informações em tempo real, de forma rápida e segura.

## 2.3 <sup>4</sup> Variedade

As fontes de dados são muito variadas, o que aumenta a complexidade das análises. Essas fontes podem ser desde transações comerciais, redes sociais, radares, sensores, telemóveis, computadores, sites de reclamações, “gostos” nas redes sociais, coordenadas de GPS, drones, etc. Além das variedades das fontes, existem também os diferentes formatos e tipos de dados, desde estruturados como dados numéricos em bancos de dados tradicionais, dados não estruturados como textos, e-mails, vídeos, áudios e dados de transações financeiras e dados semi-estruturados como JSON ou XML. <sup>4</sup> Um dos maiores desafios do *Big Data* é gerir todos esses diferentes tipos de dados e fontes ao mesmo tempo.

## <sup>11</sup> 2.4 Veracidade

A veracidade está ligada diretamente ao quanto uma informação é verídica. É a necessidade de garantir que os dados obtidos sejam autênticos. Isso aplica-se às fontes de dados que precisam de ser confiáveis e aos dados e informações propriamente ditos.

## 2.5 Valor

Refere-se à capacidade de transformar os dados em valor. O objetivo final do *Big Data* é extrair informações úteis que possam levar a decisões informadas e melhorias nos negócios. Transformar grandes volumes de dados em valor tangível é um dos principais objetivos do *Big Data*. [4]

# 3 COMO INTERPRETAR DADOS

<sup>10</sup> Qualquer informação pode ser definida em dados, mas existem diferentes formatos que podem facilitar, dificultar ou direcionar a análise feita pelas empresas. Os dados são divididos em três tipos: estruturados, semi-estruturados e não estruturados.

## <sup>13</sup> 3.1 Dados Estruturados

São organizados em linhas e colunas, em formato de tabela, e são encontrados em banco de dados, sendo muito eficientes quanto à recuperação e processamento. Folhas de cálculo e SQL (linguagem dos Bancos de Dados) são próprios para acesso e manipulação destes dados. São <sup>5</sup> facilmente reconhecidos pelos bancos e facilitam a análise e processamento.

5 Estes são derivados das interações entre pessoas e máquinas, como aplicações na net e redes sociais. Um exemplo são dados resultantes do comportamento dos utilizadores na web, também conhecidos como logs. É uma mistura de texto, imagens e dados estruturados como formulários ou informações transacionais. O avanço digital transformou muitos desses dados, trazendo ainda mais formatos para os considerados estruturados.

### 3.2 Dados Semi-Estruturados

Estes dados possuem uma estrutura flexível, o que significa que não seguem um esquema fixo, mas possuem elementos de identificação que facilitam a organização e a pesquisa. Exemplos comuns de dados semi-estruturados incluem arquivos XML, JSON e HTML. A classificação e exploração de dados semi-estruturados em *Big Data* requer técnicas de análise de dados específicas, como a utilização de consultas especializadas e algoritmos de análise de documentos. São usados por exemplo nas configurações de aplicações e perfis de utilizadores que muitas vezes são armazenados em formatos semi-estruturados devido à flexibilidade que oferecem. Os dados semi-estruturados desempenham um papel crucial em muitos contextos modernos devido à sua capacidade de acomodar a diversidade e a mudança nas informações.

### 3.3 Dados Não Estruturados

Ao contrário dos dados estruturados, os dados não estruturados não possuem um formato pré-definido e não podem ser facilmente organizados em tabelas e colunas. Estes dados derivam de diferentes fontes, como emails, documentos de texto, áudio, vídeo, sensores, entre outros. A classificação e exploração de dados não estruturados em *Big Data* pode ser um desafio, uma vez que exigem técnicas avançadas de processamento de linguagem natural, reconhecimento de padrões e machine learning. No entanto, esses dados podem fornecer informações valiosas para a tomada de decisões e análise de tendências. Alguns exemplos são informações visuais, como fotos, vídeos e gráficos, arquivos de áudio, como gravações de voz, podcasts e músicas e atividades em redes sociais, como gostos, partilhas e comentários. [13]

## 4 DADO EM MOVIMENTO OU EM REPOUSO

Os termos "dados em movimento" (*data in motion*) e "dados em repouso" (*data at rest*) são conceitos que descrevem o estado ou a condição dos dados em diferentes momentos do seu ciclo de vida, especialmente em ambientes de tecnologia da informação e gerenciamento de dados.

### 4.1 Dados em Movimento (*data in motion*)

Refere-se a dados que estão em trânsito, sendo transmitidos ou movendo-se entre sistemas. Exemplos incluem dados sendo transferidos pela rede, streams de dados em tempo real, mensagens em um sistema de mensagens, entre outros. A segurança e a integridade dos dados em movimento são preocupações fundamentais, e técnicas como criptografia e protocolos de comunicação seguros são frequentemente aplicadas. Medidas de segurança incluem o uso de protocolos seguros de comunicação, como HTTPS, e a criptografia dos dados durante a transmissão.

### 4.2 Dados em Repouso (*data at rest*)

Refere-se a dados que estão armazenados e não estão em movimento. Estes dados estão em repouso em dispositivos de armazenamento, como discos rígidos, SSDs, fitas magnéticas ou em bancos de dados. A segurança dos dados em repouso é uma preocupação crítica. Isso envolve práticas como a criptografia de dados armazenados e o controlo de acesso aos sistemas de armazenamento. Um exemplo são as cópias de dados feitas para fins de backup e recuperação, seja em

dispositivos físicos ou em serviços de backup na nuvem. A criptografia de dados armazenados, a implementação de firewalls e a gestão adequada de credenciais são práticas comuns para garantir a segurança dos dados em repouso. [14]

## 5 APLICAÇÕES DO BIG DATA EM DIFERENTES SETORES

### 5.1 Comportamento do usuário

O comportamento do usuário pode ser rastreado e recolhido com o apoio do *Big Data* a partir do uso de sistemas de *analytics*, é possível intuir determinadas ações do usuário de acordo com esse comportamento prévio. Com base em históricos de compras, em interações e em sessões de atendimento, o sistema pode prever, por exemplo, a possibilidade de um cliente pedir o cancelamento dos serviços contratados. [1]

### 5.2 Detecção de fraudes

O *Big Data* pode rastrear as informações, identificar padrões que não são normais, apontar alterações e, ainda, mapear por onde as mudanças foram feitas. Uma das técnicas mais utilizadas para garantir a segurança de dados é o Machine Learning: estes modelos podem aprender a partir de dados históricos e em tempo real a identificar comportamentos fraudulentos com base em características específicas. As máquinas são ensinadas a ler padrões e apontam quando algo de diferente acontece, alertando os responsáveis. [16]

### 5.3 Assistentes virtuais

O *Big Data* desempenha um papel significativo no desenvolvimento e aprimoramento de assistentes virtuais. São guardados dados de interações com o usuário para análise de sistemas analíticos que geram um resultado capaz de conversar e interagir de maneira espontânea. Os assistentes virtuais são baseados em machine learning e utilizam grandes conjuntos de dados para melhorar a compreensão da linguagem natural. O *Big Data* é fundamental para treinar modelos de PLN (Processamento de Linguagem Natural) em grandes volumes de dados textuais, permitindo que assistentes virtuais compreendam nuances linguísticas, contextos e variações na forma como as pessoas se expressam.

### 5.4 Saúde

O *Big Data* na saúde é um processo que consiste na extração de informações relevantes a partir de um volume elevado de dados, usualmente associados a procedimentos médicos e pacientes. Pode avaliar centenas de dados de pacientes e encontrar um padrão de diagnóstico, que agiliza o atendimento de outros indivíduos com o mesmo perfil. Também permite a redução de custos e tempo na avaliação de propostas, porque torna-se possível realizar diagnósticos mais precisos e com menos hipóteses de erros, graças à concentração de informações e identificação de padrões. Atualmente, já existem tecnologias que utilizam a inteligência artificial para avaliar relatórios e traçar um diagnóstico com base em bancos de dados. [15]

## 6 BIG DATA E INTERNET DAS COISAS (IOT)

A Internet das Coisas ou *Internet of Things* (IoT) refere-se à possibilidade de ligar diferentes dispositivos à rede mundial de computadores. O objetivo principal é promover uma integração entre tudo o que é útil no nosso dia a dia, a fim de proporcionar mais eficiência e utilidade.

## 6.1 <sup>1</sup> Otimização das relações de consumo

Há alguns anos, o consumidor comprava um produto ou serviço com base apenas nas informações do vendedor. Agora, a realidade é outra. O cliente tem a opção de ir até a loja física ver um produto, enquanto pesquisa sobre o mesmo item na internet, faz uma comparação de preços e verifica a avaliação de compradores anteriores, tudo isso de forma simultânea.

Diante disso, as empresas que desejam expandir os seus negócios precisam de conhecer bem os seus consumidores, entender quais são as suas dores e necessidades. Sobretudo, saber aproveitar essas informações disponíveis a partir do *Big Data*, para usar isso como um diferencial para a sua marca.

## 6.2 Melhora eficiência

O principal objetivo da Internet das Coisas é criar alternativas eficientes que vão facilitar as atividades do nosso dia a dia. Um exemplo disso são os serviços prestados pela Uber. Por meio da união do *Big Data* com a IoT, o motorista tem acesso a um sistema que permite obter informações referentes ao trajeto que será percorrido, como a distância e o tempo necessário para a deslocação. [9]

## 7 MACHINE LEARNING <sup>14</sup> BIG DATA

Machine learning é um método de analisar dados e informações de tal forma que os sistemas aprendam com eles e evoluam por conta própria, eliminando ou reduzindo a necessidade de intervenção humana.

<sup>6</sup> Machine learning e big data têm um relacionamento crucial dentro dos processos de tecnologia. Por si só, essas ferramentas já são grandes avanços para o mundo tecnológico. Mas, quando utilizadas em conjunto, elas podem oferecer resultados ainda melhores. [7]

### 7.1 Como o Big Data impulsiona algoritmos de machine learning

<sup>22</sup> Um algoritmo de machine learning pode perfeitamente funcionar com uma base de dados menor. Mas quando ele é combinado com o big data, <sup>6</sup> os resultados são maximizados. Um modelo de machine learning aprende muito mais e mais rápido quando é alimentado por um grande e variado volume de dados e informações.

Deste modo, o machine learning pode encontrar, em big data, padrões e anomalias que podem solucionar problemas e até criar novos *insights*, permitindo que tecnologias e empresas se desenvolvam. Ou seja, graças ao volume de dados e à rapidez com que eles chegam, as ações a serem determinadas por machine learning tornam-se mais precisas e relevantes. [8]

## 8 <sup>25</sup> PROCESSAMENTO DE LINGUAGEM NATURAL (PLN) EM BIG DATA

As tecnologias de <sup>7</sup> processamento de Linguagem Natural (PLN) fazem parte do campo da ciência da computação que engloba a inteligência artificial e a computação cognitiva, cujo foco está na interação entre os computadores e a linguagem humana.

Atualmente, a maior parte <sup>7</sup> dos dados totais produzidos no mundo estão disponíveis na sua forma bruta sem qualquer tipo de classificação. Exemplos incluem informações de pessoas, transações comerciais, processos jurídicos, entre outros. Embora o uso humano seja linguístico, ambíguo e desestruturado para ser interpretado por computadores, com a ajuda do PLN, esses dados não estruturados podem ser aproveitados para padrões evolutivos dentro de dados para conhecer melhor a informação contida nos dados. [6]

## 9 SEGURANÇA EM BIG DATA

2 segurança de base de dados são os processos, ferramentas e controlos que protegem as bases de dados contra ameaças acidentais e intencionais. O objetivo da segurança da base de dados é proteger os dados confidenciais e manter a confidencialidade, a disponibilidade e a integridade da base de dados. Além de proteger os dados na base de dados, a segurança da base de dados protege o sistema de gestão de bases de dados e aplicações, sistemas, servidores físicos e virtuais associados. 2 Para alcançar o nível mais elevado de segurança da base de dados, as organizações precisam de várias camadas de proteção de dados. Se uma camada de 2 proteção falhar, existe outra para impedir imediatamente o ataque.

### 9.1 2 Segurança da rede

As firewalls servem como a primeira linha de defesa na segurança da base de dados. Uma firewall é um separador ou um restritor do tráfego de rede, que pode ser configurado para impor a política de segurança de dados de uma organização. Se utilizar uma firewall, irá aumentar a segurança ao nível do sistema operativo ao fornecer um ponto de redução onde as suas medidas de segurança podem ser focadas.

### 9.2 Gestão de acesso

A autenticação é o processo de provar que o utilizador é quem afirma ser ao inserir o ID de utilizador e a palavra-passe corretos. Algumas soluções de segurança permitem aos administradores gerir as identidades e as permissões dos utilizadores da base de dados. 2 controlo de acesso é gerido pelo administrador de sistema que atribui permissões a um utilizador numa base de dados.

### 9.3 2 Proteção de informações

A encriptação de dados protege os dados confidenciais ao convertê-los num formato alternativo para que apenas as partes pretendidas os possam decifrar para o formato original e aceder aos dados. Embora a encriptação não resolva problemas de controlo de acesso, melhora a segurança ao limitar a perda de dados quando os controlos de acesso são ignorados. Por exemplo, se o computador anfitrião da base de dados estiver mal configurado e um utilizador mal intencionado obtiver dados confidenciais, 2 essas informações roubadas poderão não ser úteis se estiverem encriptadas.

A recuperação e os dados da cópia de segurança da base de dados são essenciais para proteger informações. Este processo envolve fazer cópias de segurança da base de dados e dos ficheiros de registo regularmente e armazenar as cópias numa localização segura. A cópia de segurança e o ficheiro estão disponíveis para restaurar a base de dados em caso de falha de segurança. [11]

## 10 GESTÃO 1 DE DADOS

Entende-se por gestão de dados, o conjunto de políticas, processos e estruturas organizacionais que apoiam a gestão 1 de dados corporativos. É através disto que as empresas gerem os dados, a sua disponibilidade, a sua utilização e a segurança das informações que dispõem.

1 Atualmente, as empresas estão a lidar com um grande número de novos dados, e nem sempre sabem qual é o caminho mais adequado para os gerir. Com a gestão de dados, é possível determinar os ambientes de big data apropriados para o acesso e armazenamento desses dados, além de fomentar uma arquitetura de dados, que possa controlar as fontes, integrá-las e disponibilizá-las aos interessados.

### 10.1 Principais objetivos da Gestão de Dados

- 1. Aprimoramento do processo de tomada de decisões, possibilitando que as decisões sejam tomadas com base em informações mais fidedignas, seguras e assertivas.
- Maior proteção das informações de investidores, pois quanto maior for o controle sobre os dados, maior será a segurança dos mesmos.
- Processos mais eficientes, pois com os padrões e a repetição proposta pela gestão de dados, os processos passam a ser mais fáceis e rápidos.
- Diminuição de custos, pois, através de uma gestão bem planeada e executada, é possível gerir o trabalho desenvolvido na área, aumentar a produtividade das equipas que utilizarão os dados e reduzir o desperdício de tempo e investimento.

### 10.2 Benefícios da Gestão de Dados

- **Maior e melhor segurança, privacidade e conformidade dos dados:** Com uma boa gestão de dados, empresas podem realizar a identificação e o tratamento diferenciado de dados pessoais, através da utilização de metadados e podendo instituir acesso restrito e controlado, que trará maior segurança aos dados.
- 1. **Maior qualidade dos dados:** Dados de baixa qualidade, implicam tomadas de decisões erradas. De nada adianta possuir boas ferramentas de negócios, se os dados que as alimentam não são bons. Muitos dados são extraídos de diversas fontes, e uma boa gestão de dados, tem o poder de conectar essas informações, através de sistemas capazes de identificar relações significativas e obter o máximo de dados possíveis, garantindo que dados importantes não fiquem para trás.
- **Automação acelerada:** Através de práticas e ferramentas que atuam na manutenção, na organização e na qualidade dos dados, os tempos de análise de dados podem automatizar tarefas e processos específicos com algoritmos de machine learning. Permitindo assim, que os dados de clientes possam ser alimentados em modelos.
- 1. Uma boa gestão de dados contribui significativamente para que os dados dos clientes sejam precisos e protegidos, e para que as equipas possam utilizar esses dados para realizar vendas mais direcionadas e assim, colaborar para o crescimento da empresa.

[10]

## 11 BIG DATA E BLOCKCHAIN

Blockchain é uma tecnologia para o registo digital de transações. Trata-se de uma tecnologia que agrupa um conjunto de informações que se conectam por meio de criptografia. Assim, transações financeiras e outras operações podem ser feitas de forma segura. Blockchain é importante para garantir que ninguém consiga efetuar fraudes, tornando cada moeda rastreável desde o momento de sua criação.

A grande inovação da Blockchain foi armazenar os dados de forma sequencial, porém sem a necessidade de uma entidade a coordenar o processo. Os próprios utilizadores da rede conseguem verificar de forma simples e praticamente sem custo se tudo está a ser cumprido. Ao contrário das redes privadas, a Blockchain permite que os dados sejam compartilhados por todos, sem necessidade de permissão. [2]



### 11.1 Como o Blockchain pode ajudar o Big Data

- **Compartilhamento de dados:** Cientistas e analistas de dados usam dados derivados de relatórios, análises e estudos. Esses dados são armazenados em blockchain e permitem o acesso de mais pessoas ao mesmo tempo. Além disso, os resultados da análise de dados derivados e compartilhados em Blockchain podem ser monetizados pelos cientistas de dados.
- **Análise de dados em tempo real:** A tecnologia Blockchain fornece análise de dados em tempo real com a máxima precisão. Em todas as transações de criptomoedas, um contrato inteligente está envolvido. O contrato inteligente contém os dados do transferidor, destinatário e carimbos da data/hora da transação. Isso pode ajudar a identificar transações suspeitas e fraudulentas. Os bancos também podem implementar a tecnologia blockchain para análise de dados em tempo real e observar melhorias significativas na tomada de decisões.
- **Previsão de dados:** O blockchain obtém dados estruturados de fontes como contas de usuário e dispositivos. Esses dados podem ajudar os cientistas de dados a prever resultados relacionados a clientes e empresas. Pesquisa de mercado, investimentos empresariais e produção industrial podem ser beneficiados com os dados previstos. Os dados ajudam ainda a estudar padrões e tendências em cada setor.

[3]

## 12 PRIVACIDADE E ÉTICA EM BIG DATA

3 análise de *Big Data* levanta uma série de questões éticas, especialmente quando as empresas começam a monetizar os seus dados externamente para finalidades diferentes daquelas para as quais os dados foram inicialmente recolhidos. A escala e a facilidade com que as análises podem ser conduzidas hoje mudam completamente a estrutura ética da área de dados. [12]

### 12.3 Os dados e a identidade de clientes devem permanecer privados

Privacidade não significa sigilo, pois os dados privados podem precisar de ser auditados com base em requisitos legais. Porém, os dados privados obtidos de uma pessoa com o seu consentimento não devem ser expostos para uso por outras empresas ou indivíduos com quaisquer vestígios da sua identidade. As identidades não podem ser expostas no universo virtual por isso, os gestores criam diretrizes, limitações e políticas internas que sejam aplicadas.

### 12.3 Informações privadas compartilhadas devem ser confidenciais

A confidencialidade dos dados é indispensável para impedir a ocorrência de crimes virtuais e fraudes que são frequentes no mundo. Todas as informações relacionadas aos clientes e fornecidas remotamente precisam permanecer em sigilo absoluto. A ética no uso dos dados dos utilizadores exige garantir a proteção da privacidade e confidencialidade das informações pessoais. Isso envolve implementar medidas de segurança adequadas para evitar acesso não autorizado, uso indevido ou divulgação dos dados recolhidos. [17]

### 12.3 Clientes devem ter uma visão transparente do processo

As empresas devem garantir que os utilizadores saibam que os seus dados estão a ser guardados e como serão usados. Isso envolve uma obrigação de transparência, ou seja, as empresas devem ser claras sobre o que estão a guardar e porque. As pessoas na posição de cliente devem ver claramente como os dados estão a ser usados ou vendidos. A gestão do fluxo das suas informações privadas deve ser feito por meio de sistemas analíticos de terceiros.

## 12.4 Impacto Social e Discriminação

Os algoritmos de Machine Learning podem absorver tendências inconscientes numa população e amplificá-las por meio de amostras de treinamento. A ética no uso de dados dos utilizadores na inteligência artificial envolve evitar a introdução de preconceitos e discriminação nos modelos e algoritmos. É fundamental garantir que as decisões e recomendações criadas pela inteligência artificial sejam imparciais, equitativas e livres de preconceitos. [5]

## 13 CONCLUSÃO

A ascensão do Big Data marca uma revolução significativa na forma como lidamos com informações nas nossas sociedades modernas. Ao longo deste trabalho, explorámos as diversas facetas dessa revolução, desde as tecnologias fundamentais que possibilitam o processamento massivo de dados até as aplicações práticas que transformam indústrias inteiras.

O Big Data não é apenas sobre a quantidade massiva de informações, mas também sobre como usamos esse conhecimento para melhorar a sociedade, impulsionar a inovação e abordar os desafios complexos que enfrentamos.

Em conclusão, o Big Data não é apenas uma ferramenta tecnológica, é uma transformação fundamental na maneira como entendemos e interagimos com o vasto oceano de informações à nossa volta. À medida que continuamos, é crucial manter um equilíbrio entre inovação e responsabilidade, garantindo que o impacto do Big Data seja construtivo, ético e benéfico para a humanidade.

## REFERENCES

- [1] Gabriel Coimbra. As aplicações de big data. Disponível em: [https://www.falandoti.com/as-aplicacoes-do-big-data/, year="2022"](https://www.falandoti.com/as-aplicacoes-do-big-data/, year=).
- [2] Coinext. O que é blockchain? aprenda tudo sobre essa tecnologia. Disponível em: <https://coinext.com.br/o-que-e-blockchain>, 2022.
- [3] Express Computer. Blockchain and big data analytics – transforming technology and beyond. Disponível em: <https://www.expresscomputer.in/news/blockchain-and-big-data-analytics-transforming-technology-and-beyond/89274/>, 2022.
- [4] Editor Rock Content. Big data: entenda o conceito, suas aplicações em diferentes contextos e como impacta as iniciativas de marketing. Disponível em: <https://rockcontent.com/br/blog/big-data/>, 2021.
- [5] Crawly. Precisamos falar sobre ética na coleta de dados. Disponível em: <https://www.crawly.com.br/blog/precisamos-falar-sobre-etica-na-coleta-de-dados>, 2023.
- [6] Santo Digital. Pln e big data: o papel do processamento de linguagem natural na big data. Disponível em: <https://www.jusbrasil.com.br/artigo/7-n-e-big-data-o-papel-do-processamento-de-linguagem-natural-na-big-data/535705948>, 2016.
- [7] Santo Digital. Machine learning ajuda as empresas na prevenção de fraudes. entenda como! Disponível em: <https://santodigital.com.br/machine-learning-ajuda-empresas-na-prevencao-de-fraudes-entenda-como/>, 2022.
- [8] Gatefy. Como machine learning e big data se relacionam. Disponível em: <https://gatefy.com/pt-br/blog/como-machine-learning-e-big-data-se-relacionam/>, 2021.
- [9] Wagner Hörlle. Big data e internet das coisas: como se relacionam. Disponível em: <https://blog.csptecnologia.com/big-data-e-internet-das-coisas/>, 2021.
- [10] Wagner Hörlle. Governança de dados que você precisa saber! Disponível em: <https://blog.csptecnologia.com/governanca-de-dados/>, 2021.
- [11] Microsoft. O que é a segurança da base de dados? Disponível em: <https://azure.microsoft.com/pt-pt/resources/cloud-computing-dictionary/what-is-database-security>, 2022.
- [12] MULTIEDRO. Ética de dados: saiba mais sobre o tema e sua importância. Disponível em: <https://blog.multiedro.com.br/etica-de-dados/>, 2022.
- [13] José Antonio Ribeiro Neto. Quais são os tipos de dados da era big data? Disponível em: <https://medium.com/xnewdata-portugal/quais-s>
- [14] José Antonio Ribeiro Neto. Classificação dos dados. Disponível em: <https://medium.com/xnewdata-portugal/cap>
- [15] NeuralMed. Big data na saúde: Por que a importância de olhar para esse universo. Disponível em: <https://www.neuralmed.ai/blog/big-data-na-saude>, 2022.
- [16] Tail. Big data no combate à fraude. Disponível em: <https://blog.tail.digital/big-data-no-combate-a-fraude/>, 2021.
- [17] Tera. 5 princípios de ética em big data. Disponível em: <https://medium.com/somos-tera/5-principios-de-etica-em-big-data-4cada3d86687>, 2020.

## ● 58% geral de similaridade

As principais fontes encontradas nos seguintes bancos de dados:

- 55% Banco de dados da Internet
- Banco de dados do Crossref
- 15% Banco de dados de trabalhos enviados
- 6% Banco de dados de publicações
- Banco de dados de conteúdo publicado no Crossref

### PRINCIPAIS FONTES

As fontes com o maior número de correspondências no envio. Fontes sobrepostas não serão exibidas.

1	<b>blog.csptecnologia.com</b> Internet	13%
2	<b>azure.microsoft.com</b> Internet	9%
3	<b>medium.com</b> Internet	4%
4	<b>pt.wikipedia.org</b> Internet	4%
5	<b>rockcontent.com</b> Internet	4%
6	<b>gatefy.com</b> Internet	3%
7	<b>sajdigital.com.br</b> Internet	3%
8	<b>neuralmed.ai</b> Internet	3%

9	<b>export.arxiv.org</b> Internet	2%
10	<b>capsulati.com.br</b> Internet	2%
11	<b>IPS Instituto Politécnico de Setubal on 2023-12-29</b> Submitted works	2%
12	<b>Universidade Estadual Paulista on 2022-10-04</b> Submitted works	1%
13	<b>vdoc.pub</b> Internet	1%
14	<b>trabalhosgratuitos.com</b> Internet	<1%
15	<b>ESPM - Escola Superior de Propaganda e Marketing on 2023-04-01</b> Submitted works	<1%
16	<b>1library.org</b> Internet	<1%
17	<b>IPS Instituto Politécnico de Setubal on 2023-12-29</b> Submitted works	<1%
18	<b>jornalcontabil.com.br</b> Internet	<1%
19	<b>santodigital.com.br</b> Internet	<1%
20	<b>Sydney Institute of Technology and Commerce on 2023-02-08</b> Submitted works	<1%

21	<b>repositorio.animaeducacao.com.br</b>	<1%
	Internet	
22	<b>sga.com.br</b>	<1%
	Internet	
23	<b>tandfonline.com</b>	<1%
	Internet	
24	<b>European University on 2022-11-23</b>	<1%
	Submitted works	
25	<b>crb8.org.br</b>	<1%
	Internet	
26	<b>passeidireto.com</b>	<1%
	Internet	
27	<b>pt.wiki2.wiki</b>	<1%
	Internet	
28	<b>ESPM - Escola Superior de Propaganda e Marketing on 2023-04-01</b>	<1%
	Submitted works	
29	<b>fw.uri.br</b>	<1%
	Internet	
30	<b>Guilherme Pereira Bento Garcia. "Análise morfométrica e identificação ...</b>	<1%
	Crossref posted content	
31	<b>Universidade de Brasília, UnB on 2022-12-13</b>	<1%
	Submitted works	
32	<b>hdl.handle.net</b>	<1%
	Internet	

33	<b>link.springer.com</b>	Internet	<1%
34	<b>atech.com.br</b>	Internet	<1%
35	<b>nucleodoconhecimento.com.br</b>	Internet	<1%
36	<b>seidor.com</b>	Internet	<1%
37	<b>Nuno Moutinho. "Meios emergentes de comunicação e educação num..."</b>	Publication	<1%
38	<b>Tatiana Kimura Kodama, Fernando José Gómez Paredes, Moacir Godin...</b>	Crossref	<1%