

SENAI AVAK BEDOUIAN

DAVI DE CARVALHO BLASIOLI

EDUARDA MASSON COSTA

JOÃO PEDRO SILVA LIMA

STHEFANY LARISSA CHIODEROLI

BIG DATA E ANÁLISE DE DADOS

Projeto e pesquisa

BIRIGUI – SP

2024

DAVI DE CARVALHO BLASIOLI
EDUARDA MASSON COSTA
JOÃO PEDRO SILVA LIMA
STHEFANY LARISSA CHIODEROLI

BIG DATA E ANÁLISE DE DADOS

Projeto e pesquisa

Projeto baseado no curso de Desenvolvimento de Sistemas com ênfase em Big Data e Análise de Dados na Instituição de Ensino e Aprendizagem SENAI Avak Bedouian. Pesquisa apresentada ao curso referente, da área da tecnologia, da Instituição supracitada.

Orientador(a): Prof. Igor Cacerez e Lais Ribeiro Sinatra

BIRIGUI – SP

2024

RESUMO

Este documento tem como finalidade expor a temática relacionada a Big Data e análise de dados, destacando a importância e o impacto dessas tecnologias, não somente na era digital, como também no cotidiano da sociedade de maneira geral. Tal projeto foi solicitado por meio da Instituição SENAI, relacionado ao curso de Desenvolvimento de Sistemas. A partir dele, é possível realizar uma análise mais profunda e eficiente dos conceitos referentes ao tema e suas aplicações gerais. A pesquisa permite um aprofundamento nos estudos sobre os Dados e os mais diversos âmbitos que nisso agrega.

O termo Big Data refere-se à análise de grandes volumes de dados que são gerados a uma velocidade gradativa, cujo formatos são variados, evidentes em inúmeras fontes, como redes sociais, dispositivos móveis e transações comerciais, por exemplo. A análise de dados, por sua vez, engloba técnicas e ferramentas para extrair informações fundamentais a partir desses dados, facilitando a tomada de decisões relevantes e a identificação de padrões e tendências.

O trabalho conclui com discussões referentes a desafios e oportunidades que os tais possuem, apresentando quesitos de privacidade, segurança e exemplos reais que lidam com essas tecnologias. As implicações na comunidade dessas ferramentas são também consideradas, abordando a importância da utilização de dados de forma harmônica e ética.

Palavras-chave: Big Data, Análise de Dados, Desenvolvimento de Sistemas, Dados.

ABSTRACT

This document aims to expose the topic related to Big Data and data analysis, highlighting the importance and impact of these technologies, not only in the digital era, but also in the daily life of society in general. This project was requested through the SENAI Institution, related to the Systems Development course. From there, it is possible to carry out a deeper and more efficient analysis of the concepts relating to the topic and its general applications. The research allows for in-depth studies on Data and the most diverse areas that it brings together.

The term Big Data refers to the analysis of large volumes of data that are generated at a gradual speed, whose formats are varied, evident in numerous sources, such as social networks, mobile devices and commercial transactions, for example. Data analysis, in turn, encompasses techniques and tools to extract fundamental information from this data, facilitating relevant decision-making and the identification of patterns and trends.

The work concludes with discussions regarding the challenges and opportunities they have, presenting privacy, security issues and real examples that deal with these technologies. The community implications of these tools are also considered, addressing the importance of using data in a harmonious and ethical way.

Keywords: Big Data, Data Analysis, Systems Development, Data.

SUMÁRIO

1. INTRODUÇÃO	7
2. DESENVOLVIMENTO	8
2.1 O QUE É	8
2.2 PARA QUE SERVE	8
2.3 COMO FUNCIONA (COLETA E ARMAZENAMENTO)	9
2.4 TECNOLOGIA E FERRAMENTAS	10
2.5 PROCESSO DE ANÁLISE DE DADOS	12
2.6 APLICAÇÕES PRÁTICAS	13
2.7 IMPORTÂNCIA E IMPACTO NA SOCIEDADE	14
2.8 BENEFÍCIOS E DESAFIOS	15
2.9 EMPRESAS QUE UTILIZAM	16
2.10 DADOS ESTATÍSTICOS	18
3. CONCLUSÃO	20
4. REFERÊNCIAS BIBLIOGRÁFICAS	21

LISTA DE IMAGENS

1. FIGURA 1	19
-------------	----

1. INTRODUÇÃO

A sociedade contemporânea se encontra em desenvolvimento tecnológico gradativo. Dados apontam que há grande praticidade em colecionar dados, visto que os processos se tornaram rápidos, seguros e eficazes. Nesse contexto, é notório que estes são gerados em uma quantidade significativa diariamente, visto que a cada acesso, clique, transação e postagem em redes sociais, as informações são produzidas em abundância.

Diante dessa realidade, é presumível que interrogações surgem, tais como: quem lida com essa relevante quantidade de dados? Como é possível extrair insights pertinentes deles? Como funciona o armazenamento e coleta de tais?

Através de inúmeros questionamentos e situações que atualmente perpetuam, é necessário ressaltar que os sistemas do Big Data e Análise de Dados são sobressalentes em diversos meios, uma vez que, além de contribuírem com o armazenamento e organização, eles atuam com a finalidade de direcionar variadas companhias a partir de procedimentos de tomada de decisões, que acarretam em realizações com maior estratégia e assertividade. Ademais, a Análise de Dados, que possui grande evidência na sociedade, visa extrair, por meio dos dados, informações e relevantes revelações.

Com o avanço tecnológico em um mundo habituado, os dados se baseiam em três básicos e essenciais conceitos: volume, variedade e velocidade, que se constituem à quantidade de dados (tendo a possibilidade de ser grande e difícil de ser processada através de métodos tradicionais), à diversidade de formatos e tipos de dados (podendo variar desde textos até sensores) e, por fim, à agilidade para os dados serem processados em tempo real, respectivamente.

Tendo em consideração os aspectos supracitados, é de extrema importância relacionar estes com o célebre Big Data, que cresce gradativamente. Este é fundamental para a contemporaneidade, portanto, é evidente em inúmeros setores, como saúde, marketing, negócios, finanças, entre outros. Tais meios a utilizam a fim de desfrutarem das informações significativas e usufruir vantagens e diferenciais, contribuindo com os clientes, com a personalização de produtos e serviços, identificação de tendências de mercado e antecipação de demandas posteriores. Entretanto, para lidar com o Big Data é necessário a realização de análises eficazes com o uso de ferramentas e técnicas exclusivas (tendo como exemplo a Inteligência artificial, tão notória na atualidade), abrangendo medidas de armazenamento e processamentos dos dados em bancos de dados ou nuvem.

Para concluir, é possível notar que tais conceitos permitem transformar dados em conhecimento, impulsionando a inovação e o sucesso nos negócios e no mundo digital.

2. DESENVOLVIMENTO

O Big Data e a Análise de Dados estão interligados e possuem diversos âmbitos a serem desvendados. Eles são capazes de oferecer informações de maneira rápida, segura e eficaz.

2.1 O QUE É

No início dos anos 2000, o analista Doug Laney percebeu a importância da coleta e armazenagem de dados para futuras pesquisas. Desse modo, em 2001 surgiu a Big Data, que é um termo usado para descrever grandes volumes de dados, que são coletados, armazenados, analisados e interpretados, os quais podem vir de diversas formas ou fontes, como transações bancárias, registros de atividades online, redes sociais, entre outros, que normalmente uma ferramenta tradicional não conseguiria realizar de forma eficiente e rápida.

O Big Data tem como principal função ajudar empresas nas tomadas de decisões e análises que levam a descobertas e investimentos importantes para diversas áreas de uma empresa.

O termo se iniciou com os 3 V's: **Volume**, que faz referência a enorme quantidade de dados gerados diariamente, que pode chegar a exabytes, **Velocidade** é a questão da agilidade do processamento de dados, e **Variedade** que se refere aos diferentes formatos que os dados chegam, como imagem, vídeo, texto, dados estruturados e não estruturados. E algumas abordagens atuais também são consideradas, que são a **Veracidade** que representa a confiabilidade dos dados coletados e o **Valor**, o qual se refere aos benefícios que uma empresa pode ter após a coleta de dados.

2.2 PARA QUE SERVE

A principal utilidade do Big Data é permitir que organizações e empresas extraiam informações valiosas de grandes volumes de dados para apoiar a tomada de decisões, identificar tendências, prever comportamentos e otimizar operações.

Diante disso, é notório que os gestores terão acesso a um considerável banco de dados, e ao mercado em que ele está inserido, esses dados se transformam em informações, e essas informações são usadas pela companhia

da forma que ela julgar necessário, como para entender seus principais problemas, priorizar e solucionar eles.

Além de disponibilizar esses dados para a companhia, o Big Data também ajuda a organizar eles, podendo definir quais dados são mais importantes, cruzar informações e ajudar a entender o que esses números significam.

2.3 COMO FUNCIONA (COLETA E ARMAZENAMENTO)

A funcionalidade da análise de dados da Big Data envolve completamente o processo da extração de um grande volume de dados que são extremamente complexos, o que os tornam difíceis de serem analisados por métodos tradicionais, com o objetivo de auxiliar durante as tomadas de decisões. Esse processo de análise pode ser dividido em etapas, e ele requer o uso de algumas tecnologias avançadas, algoritmos e técnicas analíticas. Alguns dos passos da funcionalidade:

- **Coletas de dados:** Essas coletas são feitas em diversas fontes, podendo ser elas:
 - a. Estruturados: Bancos de dados tradicionais, tabelas, planilhas e colunas, etc.
 - b. Não Estruturados: Textos, imagens, vídeos, etc. Esses dados podem ser coletados de diversas fontes, como de bancos de dados, posts em redes sociais, navegação na web e até por localizações do GPS.
- **Armazenamento de dados:** Eles normalmente são feitos com dois tipos de armazenamento, O armazenamento distribuído, que é usado devido ao grande número de dados armazenados, e o armazenamento em nuvem que melhora a flexibilidade no armazenamento dos dados, sendo usado por diversas organizações, como a Amazon AWS e a Google Cloud.
- **Pré-processamento de dados:** O processo é feito com o objetivo de diminuir a quantidade de dados desnecessários, um exemplo disso é quando os dados duplicados são apagados.
- **Análise de dados:** Existem 4 formas de analisar os dados nessa parte, são elas:
 - a. Análise descritiva: Ela analisa coisas que já aconteceram de acordo com cálculos padronizados, como por exemplo quantas pessoas clicaram em um anúncio;

- b. Análise diagnóstica: Assim como o nome já diz ela serve para ajudar a entender o porquê algo aconteceu, identificando padrões e situações, por exemplo, o motivo do anúncio não receber tantas visitas;
- c. Análise Preditiva: Ela serve para analisar acontecimentos importantes do passado e com base nisso ajudar a tomar boas decisões em ações futuras, mas mesmo com isso ainda não é possível ver de fato quais vão ser as consequências dessas ações, só é possível tirar uma base;
- d. Análise Prescritiva: Essa análise é basicamente um segundo passo da Preditiva, pois ela serve para mostrar as consequências e te ajudar a “escolher” a melhor delas e ter mais dedicação quanto a ela.

2.4 TECNOLOGIA E FERRAMENTAS

Existem inúmeras tecnologias, ferramentas e arquiteturas englobada ao Big Data e a Análise de Dados, dentre elas estão:

- **Hadoop**: Apache Hadoop é um dos frameworks mais utilizados para o processamento distribuído de grandes conjuntos de dados. Ele possui dois componentes principais: **HDFS** (Hadoop Distributed File System) - sistema de arquivos distribuído que armazena grandes volumes de dados em clusters de servidores – e **MapReduce** - um modelo de processamento paralelo que divide tarefas em pequenos blocos que são processados em diferentes nós no cluster.
- **Spark**: Apache Spark é um framework de processamento de dados em memória que permite a análise de Big Data de maneira muito mais rápida que o Hadoop MapReduce, principalmente para dados que requerem operações iterativas. Ele suporta diversas APIs para Python, Java, Scala e R, além de ter uma forte integração com sistemas de armazenamento distribuído, como HDFS e Apache Cassandra.
- **NoSQL**: Os bancos de dados NoSQL são projetados para lidar com grandes volumes de dados não estruturados ou semiestruturados. Eles são flexíveis e escaláveis, permitindo o armazenamento eficiente de dados em tempo real.
- **ElasticSearch**: ElasticSearch é um motor de busca e análise baseado em texto que permite pesquisas rápidas em grandes volumes de dados. Ele é amplamente usado em conjunto com Kibana (uma ferramenta de visualização) para criar dashboards e relatórios interativos.
- **Kafka**: Apache Kafka é uma plataforma de streaming de eventos distribuída usada para construir pipelines de dados em tempo real. Ele

permite a coleta, processamento e análise de fluxos de dados em tempo real.

Outros exemplos de ferramentas são:

- **Tableau:** O Tableau é uma das ferramentas de visualização de dados mais populares, com uma interface de arrastar e soltar que facilita a criação de dashboards interativos a partir de grandes conjuntos de dados. Ele pode se conectar a uma variedade de fontes de dados, incluindo Hadoop, Spark, e bancos de dados NoSQL.
- **Power BI:** O Power BI, da Microsoft, é uma ferramenta de análise e visualização de dados que se conecta a várias fontes de dados e permite a criação de relatórios e dashboards. É uma opção popular no mundo dos negócios.
- **D3.js (JavaScript):** D3.js é uma biblioteca JavaScript poderosa para criar visualizações de dados dinâmicas e interativas diretamente na web. Embora exija mais conhecimento técnico do que ferramentas como Tableau ou Power BI, oferece total controle sobre como os dados são apresentados.
- **RapidMiner:** RapidMiner é uma plataforma de ciência de dados que oferece uma interface amigável para executar análises complexas. É amplamente utilizada para aprendizado de máquina, mineração de dados e análise preditiva.
- **DataRobot:** DataRobot é uma plataforma de machine learning automatizada que ajuda a criar e implementar modelos de aprendizado de máquina rapidamente, sem a necessidade de experiência profunda em programação.

Há também a Análise de dados com Aprendizado de Máquina (Machine Learning), são elas:

- **TensorFlow:** TensorFlow, desenvolvido pelo Google, é uma biblioteca de código aberto para aprendizado de máquina. Ele é amplamente utilizado em análises de dados complexas, como reconhecimento de padrões e previsão de séries temporais.
- **Scikit-Learn:** Scikit-learn é uma biblioteca de aprendizado de máquina para Python que inclui uma ampla gama de algoritmos como regressão, classificação, clustering (técnica de aprendizado de máquina não supervisionada que junta dados em conjuntos), e processamento de dados.
- **PyTorch:** PyTorch é outra biblioteca popular de deep learning e machine learning, conhecida pela sua flexibilidade e uso em pesquisas

de ponta, especialmente em processamento de linguagem natural (NLP) e visão computacional.

O armazenamento de dados em Big Data pode ser visto em:

- **Amazon S3:** Amazon S3 (Simple Storage Service) é uma plataforma de armazenamento na nuvem altamente escalável, usada por empresas para armazenar grandes volumes de dados. Ele é ideal para armazenar dados não estruturados e arquivos de mídia.
- **Google BigQuery:** BigQuery é um banco de dados analítico na nuvem da Google, projetado para análises rápidas em grandes volumes de dados. Ele permite consultas SQL em grandes conjuntos de dados com uma alta velocidade de processamento.
- **Azure Data Lake:** Azure Data Lake é uma solução de armazenamento na nuvem da Microsoft para grandes volumes de dados, permitindo que empresas coletem, armazenem e analisem dados em um ambiente seguro e escalável.

Ademais, o processamento em Tempo Real funciona da seguinte maneira:

- **Apache Flink:** Apache Flink é um framework de processamento de dados em tempo real. Ele se destaca no processamento de streams de dados em tempo real, com baixa latência e alta taxa de transferência.
- **Storm:** Apache Storm é uma tecnologia que permite o processamento de dados em tempo real e é bastante usado para análise de dados ao vivo em sistemas de grande escala.

2.5 PROCESSO DE ANÁLISE DE DADOS

Primeiramente, é necessário definir e localizar o problema que deseja solucionar, pois sem o objetivo pré-definido a análise de dados se torna sem sentido, ineficaz e desorientada. Após isso, ocorre a coleta dos dados, que realiza a análise de dado, sendo uma das etapas mais importantes e críticas, em que se pode buscar os dados necessários para realizar a sua análise. Tais dados podem ser obtidos de diversas fontes, como bases de dados internas, APIs, planilhas etc. Alguns tipos de dados incluem estruturados (como tabelas em banco de dados) e não estruturados (textos, imagens, vídeos etc.)

Os dados brutos geralmente contêm erros, valores ausentes, duplicados e inconsistências. Portanto, limpar e preparar os dados envolve a remoção ou correção desses problemas. Essa etapa também pode incluir a transformação dos dados, como normalização, categorização ou padronização.

Além disso, há a exploração dos dados, em que é utilizado ferramentas estatísticas e de visualização para explorar os dados. O objetivo é entender padrões, identificar tendências, detectar outliers e fazer hipóteses iniciais. Ferramentas comuns incluem, gráficos de dispersão, histogramas, tabelas dinâmicas etc. Em seguida, ocorre a modelagem, que dependendo do objetivo, pode ser dividida em Modelos Descritivos (explicam o que já aconteceu - média, mediana, variância etc.), Modelos Preditivos (tentam prever resultados futuros com base em dados históricos - modelos de regressão, algoritmos de aprendizado de máquina) e Modelos Prescritivos (recomendam ações ou estratégias - otimização).

Vale ressaltar que, técnicas avançadas como aprendizado de máquina e inteligência artificial podem ser aplicadas nesta fase para criar previsões ou identificar padrões complexos.

Após a modelagem é necessário traduzir os resultados obtidos para um formato que os tomadores de decisão possam compreender sua análise. Isso pode incluir gráficos, relatórios ou imagens). A visualização clara pode facilitar a comunicar descobertas da sua análise de forma mais eficiente.

Com os insights (percepções) extraídos dos dados, chega à etapa da tomada de decisão, onde são desenvolvidas estratégias, melhorar os processos, ou ajustar o curso das ações em andamento. Por fim, há o monitoramento e a atualização

A análise de dados é um processo contínuo. Após sua conclusão, é de extrema importância monitorar os resultados e ajustá-los conforme o necessário, conforme os dados mudam, novos desafios e problemas surgem.

2.6 APLICAÇÕES PRÁTICAS

Diante do supracitado, tais conceituações são evidentes em diversos meios, variando dependendo da maneira aplicada. O Big Data cruza informações das mais variadas fontes, tendo como exemplo o banco de dados, o qual armazena, por exemplo, os cadastros de consumidores, históricos de mensagens e interações com o cliente. Por meio dele, há o monitoramento das mídias sociais e inúmeros percursos. No âmbito do varejo, o Big Data pode ser fortemente encontrado para a melhor administração dos negócios, algumas aplicações são:

- Previsão de vendas e performance;
- Estratégias de marketing;
- Otimização de e-commerce;

- Personalização de ofertas;
- Aumento de vendas;
- Redução de custos;
- Recomendações;
- Reconhecimento dos clientes.

Tais recursos facilitam a gestão do ambiente, uma vez que, como supracitado, analisa as tendências e prevê elas, contribuindo com vendas e recomendações aos clientes (oferecendo soluções adequadas). Refletindo além do varejo, as aplicações podem ser vistas na área da saúde, com a criação de fichas eletrônicas, recebimento de alertas e previsões de necessidades, por exemplo.

Dessa maneira, um exemplo prático, muito utilizado no dia a dia, é a plataforma Google Now, que funciona como um assistente pessoal inteligente capaz de “aprender” e de se adaptar a rotina do usuário e sugerir, por exemplo, o meio de transporte disponível. Ademais, há também os aplicativos de viagem, os quais dispõem rotas com a finalidade de chegar em um determinado local, analisando os dados dos trânsitos no momento exato. A Análise de Dados contribui para essa situação exposta pois consegue analisar os feedbacks produzidos pelos clientes, direcionando o trajeto baseado na velocidade e na via.

A partir disso, conclui-se que tais tecnologias possuem muitas aplicabilidades na rotina da população, não somente em setores específicos e em empresas de grande porte, mas também em aplicativos presentes nos dispositivos, por exemplo. Contudo, vale ressaltar que o aproveitamento de dados de modo inteligente é de extrema importância, devido a obtenção de informações fundamentais e ampliação da rentabilidade dos fatores.

2.7 IMPORTÂNCIA E IMPACTO NA SOCIEDADE

A importância do Big Data está profundamente ligada à sua capacidade de transformar uma avalanche de dados em informação útil e acionável. O poder de processar grandes volumes de dados a altas velocidades tem potencial para impulsionar a inovação, otimizar processos e melhorar a tomada de decisões em uma ampla gama de setores. Vamos explorar o impacto do Big Data em algumas áreas-chave:

- **Governo e setor público:** O Big data tem um papel de extrema importância na melhoria dos serviços públicos e na gestão eficiente

das cidades. O Ministério da Justiça do Brasil, por exemplo, utiliza o Big Data para cruzar informações de pessoas envolvidas em diversos processos judiciais, permitindo ações mais efetivas contra crimes de lavagem de dinheiro.

- **Transporte e Logística:** Empresas de transporte estão utilizando o Big Data para otimizar suas rotas de entrega, economizando milhões de litros de combustível por ano e melhorando a eficiência de suas operações.
- **Indústria e Manufatura:** Muitas empresas estão usando o Big Data para otimizar sua cadeia de suprimentos e logística, garantindo que seus produtos cheguem ao mercado na hora certa, sem atrasos.
- **Varejo e Comércio Eletrônico:** Grandes varejistas, como a Amazon e a B2W Digital, estão utilizando o Big Data para entender melhor o comportamento de seus clientes e para prever as tendências de vendas, permitindo que eles mantivessem um nível ideal de estoque e evitar a perda de vendas.
- **Saúde:** Na área da saúde, o Big Data tem o potencial para revolucionar a forma como a doença é diagnosticada e tratada. Ele pode ajudar a identificar padrões e correlações em grandes conjuntos de dados de saúde, melhorando a precisão do diagnóstico e permitindo tratamentos mais personalizados.
- **Tecnologia e Mídia Social:** Plataformas de mídia social estão utilizando Big Data para personalizar a experiência do usuário, fornecendo conteúdo mais relevante e melhorando a interação do usuário com a plataforma.
- **Serviços Financeiros:** Empresas do setor financeiro usam o Big Data para identificar padrões de transações, prever possíveis inadimplências e cancelamentos de conta, e otimizar suas estratégias comerciais.

2.8 BENEFÍCIOS E DESAFIOS

O Big Data é de extrema importância para diversas áreas do mercado pois apresenta muitos benefícios, como: Eficiência operacional, que ajuda na análise dos sistemas e algum mal funcionamento, assim prevendo problemas e redução de custos, personalização que é de extrema utilidade, pois ajuda na distribuição de serviços de acordo com a preferência do cliente, inovação e previsão de tendências são essenciais para uma empresa, visto que elas ajudam a antecipar tendências futuras e ações com base nos padrões históricos, prevendo mudanças no mercado e melhorias. O Big Data também ajuda na identificação de ameaças de segurança cibernética de forma mais eficaz.

Apesar de suas diversas qualidades, o Big Data também apresenta alguns desafios, como: O volume de dados gerados, que é de muito grande, e a variedade dos dados, fazendo com que as empresas precisem investir em infraestruturas robustas, para uma melhor funcionalidade. O custo entra nesse quesito, pois é necessário investir em tecnologia, manutenção da plataforma e treinamento de pessoal, para a análise e interpretação de dados gerados é necessário um especialista em ciência de dados para melhor resultado.

2.9 EMPRESAS QUE UTILIZAM

Existem muitas empresas nacionais e internacionais que utilizam esses recursos, dentre elas estão presentes:

- **Monsanto:** aproveita a análise de dados para elaborar projetos otimizados de plantio. Os agricultores estão sempre perguntando sobre quais sementes plantar, quanto, onde e quando. Essa empresa usa a ciência dos dados para fazer recomendações prescritivas para o plantio. Eles usam os modelos matemáticos e estatísticos para planejar os melhores momentos e locais para cultivar plantas masculinas e femininas. A ideia é maximizar o rendimento e reduzir a utilização da terra. Seu algoritmo de aprendizado de máquina atinge mais de 90 bilhões de pontos de dados em dias, em vez de semanas ou meses. Isso pode reduzir e otimizar a extensão da área de plantio, inclusive.
- **Grupo Pão de Açúcar:** tem um sistema de relacionamento com o cliente voltado para a fidelização do seu público externo. Chamado de Clube Extra, o sistema tem o objetivo de não só promover a aproximação com clientes, mas também com fornecedores. De forma simples, o cliente se cadastra no programa da rede de supermercados e pode acumular pontos por meio de compras online ou em lojas físicas. Os dados originados com essa plataforma são analisados para relacionar os clientes com os produtos, com as marcas favoritas e com os mais consumidos. Operacionalmente, o sistema com a tecnologia Big Data otimiza o estoque, visto que a empresa terá conhecimento prévio sobre o quanto se deve comprar de determinado produto, tendo em vista os custos de mantê-lo.
- **Ministério da Justiça:** as instituições governamentais também estão investindo no uso de Big Data. O Ministério da Justiça do Brasil usa um banco de dados imenso, com mais de 1 bilhão de registros. Para poder analisar todos os dados, o Ministério da Justiça conta com o auxílio da tecnologia Watson da IBM, desenvolvida para coletar e processar dados em milésimos de segundos. Ele ainda utiliza o Big

Data para identificar ações ilícitas, especialmente relacionadas com lavagem de dinheiro.

- **Zara:** antes que as portas se abram diariamente em cada um dos 2.213 estabelecimentos da Zara em todo o mundo, os funcionários e os gerentes compartilham detalhes dos artigos mais vendidos do dia anterior, peças devolvidas pelos clientes, feedback dos compradores, bem como tendências que a equipe tem percebido. Usando um sistema sofisticado orientado para a tecnologia, analistas divulgam as atualizações diárias e usam-nas para pintar uma imagem precisa do que exatamente os clientes da Zara estão exigindo. Essa informação é rapidamente traduzida por uma vasta equipe de mais de 300 designers internos em projetos tangíveis que obedecem às tendências de moda, que são decentemente feitos e vendidos a preços acessíveis. E, em 21 dias, a nova peça está disponível no varejo. Enquanto a maioria das marcas de moda rápida tentam antecipar o que os clientes desejam, a Zara ganha a tendência, obtendo os comentários dos clientes. As lojas são abastecidas duas vezes por semana — com pedidos uma vez antes do fim de semana e uma vez depois. O estoque é novo e o cliente geralmente pode encontrar itens que são exclusivos. A Zara está sempre preparada para dar aos consumidores o que eles querem ou, melhor ainda, o que eles nem sabem que precisam. E, ao que parece, esse é o segredo do seu sucesso.
- **NASA:** não chega a ser uma grande surpresa, mas a Agência Espacial Norte-americana (NASA) tem diversos programas que contam com o uso de Big Data. Um exemplo é o projeto de pesquisa sobre mudanças climáticas, com análises de dados importantes coletados por meio de 16 satélites de ciências da terra da NASA para o programa de ciência climática, monitorando a qualidade do ar, os oceanos e os furacões, entre outros. Outro projeto é o supercomputador Pleiades — o supercomputador mais avançado do mundo para modelagem e simulação. Ele é um dos mais poderosos instalados no Centro de Pesquisa da NASA em Moffett Field, Califórnia, e apoia as missões da agência na exploração da ciência terrestre e espacial, aeronáutica, futuras viagens espaciais e explorações. Pleiades já foi usado para explicar a atmosfera do sol por meio de modelos numéricos avançados. Usado para modelagem e simulação, esse supercomputador ajuda os pesquisadores da NASA a aproveitar os recursos computacionais vastos e incorporar simulações em escalas espaciais.
- **Nike:** é líder mundial em várias categorias de calçados e vestuários esportivos e está investindo pesadamente em aplicativos, wearables e Big Data. A Nike está olhando além dos produtos físicos e pretende criar marcas de estilo de vida que os atletas não querem correr sem. A

empresa tem 13 linhas diferentes, em mais de 180 países. No entanto, a forma como ela segmenta e serve esses mercados é o seu diferencial real. Nike divide o mundo em empreendimentos esportivos em vez de apenas em geografia. A teoria é que as pessoas que jogam golfe, por exemplo, têm mais em comum do que as pessoas que simplesmente vivem próximas umas das outras. Essa estratégia de varejo e marketing é, em grande parte, impulsionada por Big Data. A Nike tem investido também em análise de dados demográfica para definir seus mercados de teste e impulsionar seus negócios.

2.10 DADOS ESTATÍSTICOS

Ao longo dos anos, os dados gerados têm aumentado gradativamente, devido ao crescimento de várias ferramentas inteligentes baseadas em eles, a exemplo da Inteligência Artificial e dispositivos IoT. Em proporção a isso, a velocidade de tais também se encontra neste processo. Segundo estudos de especialistas, a população gera sobre aproximadamente 2.5 quintilhões de bytes de dados por dia. Além disso, o mesmo aponta que o mercado de análise de Big Data atingirá um valor em torno de \$103 bilhões por 2027.

Na contemporaneidade, inúmeras organizações têm investido em Big Data, Análise de Dados e IA, atingido cerca de 97.2% delas. No entanto, muitas delas não têm optado por isso, em decorrência dos impedimentos gerados através da incapacidade da compreensão e acesso de tais recursos. A principal razão gira em torno do fato de que apenas 26% das empresas possuem uma cultura orientada a dados, apesar da enorme quantidade de criação de dados diária.

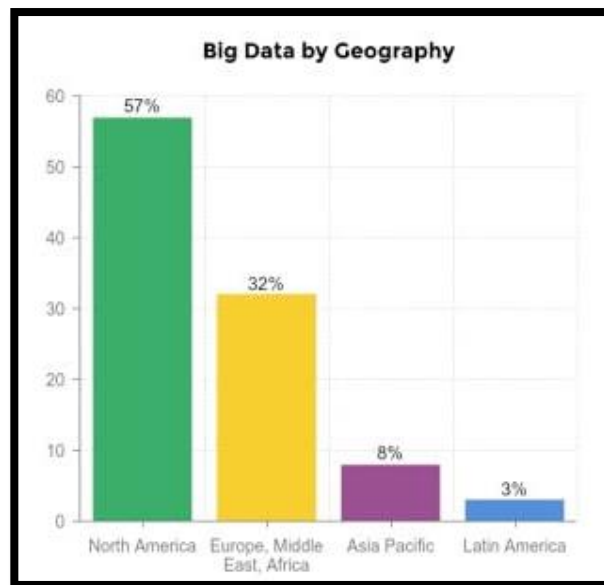
Segundo o Relatório de 2020 da MicroStrategy, foi descoberto que cerca de 94% das empresas sugerem que os dados e análises serão essenciais para o crescimento da empresa e sua transformação digital. Tal afirmação não é diferente da realidade, pois tais ambientes se superam após o investimento em tais recursos, aumentando os hábitos de consumo. A colocação analisa que as empresas que mais investem em análise de dados atualmente são hotéis, telecomunicações e varejo.

Em contradição ao fato supracitado, há inúmeras empresas que não possuem acesso a esses recursos. Além da falta de conhecimento a esse meio, muitas delas afirmam que há falta de orçamento em TI e desafios de segurança de dados.

Vale ressaltar que de acordo com pesquisas realizadas pelo International Data Corp (IDC), os Estados Unidos é o país com maior mercado de investimento em Big Data e Análise de Dados, com receitas que atingem cerca

de 88 bilhões de dólares, com crescimento gradativo a cada ano. Seguindo esse ranking, a Europa Ocidental é o segundo maior mercado geográfico para investimentos em Big Data, e o Reino Unido é o terceiro, geralmente acompanhado pelo Japão.

Figura 1 – Mercados geográficos com investimento em Big Data



Fonte: findstak (big data e estatística)

Na sociedade atual, o número de dados é tão grande e relevante que, segundo estudos de organizações, levaria cerca de 181 milhões de anos para baixar todos eles na Internet.

3. CONCLUSÃO

Em conclusão, o Big Data se consolida como uma ferramenta essencial no cenário contemporâneo, onde o volume massivo de informações gerado diariamente requer métodos eficazes de armazenamento, processamento e análise.

Através de suas características fundamentais: volume, variedade e velocidade, o Big Data permite não apenas o gerenciamento eficiente de dados, mas também a extração de insights valiosos que orientam decisões estratégicas em diversos setores, como saúde, negócios e marketing. Com o suporte de tecnologias avançadas, como a inteligência artificial, as organizações podem personalizar serviços, antecipar tendências e melhorar sua competitividade no mercado.

Dessa forma, o Big Data se estabelece como um impulsionador de inovação, agregando valor ao transformar dados em conhecimento e oferecendo um diferencial crucial para o sucesso no mundo digital.

4. REFERÊNCIAS BIBLIOGRÁFICAS

ORBY INTERNET E DESIGN. **O mundo está cada vez mais rápido e digital. E a sua empresa? - Key Access.** Disponível em: <<https://www.keyaccess.com.br/posts/detalhes/o-mundo-esta-cada-vez-mais-rapido-e-digital-e-a-sua-empresa>>. Acesso em: 18 set. 2024.

FÁBIO CRANCHI. **Introdução aos Conceitos Básicos de Big Data e Análise de Dados - GF Info - Soluções em Tecnologia.** Disponível em: <<https://www.gfinfo.com.br/introducao-aos-conceitos-basicos-de-big-data-e-analise-de-dados/>>. Acesso em: 18 set. 2024.

SEO. **7 exemplos de Big Data aplicados no varejo para aderir já.** Disponível em: <<https://www.totvs.com/blog/gestao-varejista/exemplos-de-big-data/>>.

12 aplicações de Big Data e seus benefícios. Disponível em: <<https://transformacaodigital.com/dados/aplicacoes-de-big-data-e-seus-beneficios/>>.

Estatísticas referentes a Big Data. Disponível em: <<https://findstack.com.br/resources/big-data-statistics/>>.

FERREIRA, K. **Análise de dados: o que é, como fazer e dicas básicas.** Disponível em: <<https://rockcontent.com/br/blog/analise-de-dados/#:~:text=A%20an%C3%A1lise%20de%20dados%20%C3%A9%20um%20trabalho%20dedicado%20ao%20aprofundamento>>. Acesso em: 18 set. 2024.

O que é Big Data e para que serve? Disponível em: <<https://www.zendesk.com.br/blog/o-que-e-big-data/>>.

CETAX. **Big Data: O que é, conceito e definição | Blog Cetax.** Disponível em: <<https://cetax.com.br/big-data/>>.

Leandro **Guimarães.** Disponível em: <<https://www.knowsolution.com.br/conheca-4-tipos-de-analise-de-dados-de-big-data/>>. Acesso em: 18 set. 2024.