



UNIVERSIDADE FEDERAL DO CEARÁ
PROGRAMA DE PÓS-GRADUAÇÃO EM ECONOMIA (PROFISSIONAL)
MESTRADO PROFISSIONAL EM ECONOMIA DO SETOR PÚBLICO - MESP

JOAQUIM NERSON MOURA FILHO

**BIG DATA: OPORTUNIDADES E DESAFIOS DA GESTÃO PÚBLICA LOCAL NA
CRIAÇÃO DE CIDADES MAIS INTELIGENTES**

FORTALEZA

2024

JOAQUIM NERSON MOURA FILHO

**BIG DATA: OPORTUNIDADES E DESAFIOS DA GESTÃO PÚBLICA LOCAL NA
CRIAÇÃO DE CIDADES MAIS INTELIGENTES**

Dissertação apresentada ao Programa de Pós-Graduação em Economia – CAEN, Curso de Mestrado Profissional em Economia, da Universidade Federal do Ceará – UFC, como requisito parcial para obtenção do título de Mestre em Economia. Área de Concentração: Economia do Setor Público.

Orientador: Prof. Dr. Fabrício Carneiro Linhares.

FORTALEZA

2024

Dados Internacionais de Catalogação na Publicação
Universidade Federal do Ceará
Sistema de Bibliotecas
Gerada automaticamente pelo módulo Catalog, mediante os dados fornecidos pelo(a) autor(a)

M887 Moura Filho, Joaquim Nerson.

Big data : oportunidades e desafios da gestão pública local na criação de cidades mais inteligentes /
Joaquim Nerson Moura Filho. – 2024.
111 f. : il. color.

Dissertação (mestrado) – Universidade Federal do Ceará, Faculdade de Economia, Administração,
Atuária e Contabilidade, Programa de Pós-Graduação em Economia, Fortaleza, 2024.

Orientação: Prof. Dr. Fabrício Carneiro Linhares.

1. Big Data. 2. Aprendizado de Máquina. 3. Análise de Dados. 4. Computação em Nuvem. 5. Cidades
Inteligentes. I. Título.

CDD 330

JOAQUIM NERSON MOURA FILHO

**BIG DATA: OPORTUNIDADES E DESAFIOS DA GESTÃO PÚBLICA NA CRIAÇÃO
DE CIDADES MAIS INTELIGENTES**

Dissertação apresentada ao Programa de Pós-Graduação em Economia – CAEN, Curso de Mestrado Profissional em Economia, da Universidade Federal do Ceará – UFC, como requisito parcial para obtenção do título de Mestre em Economia. Área de Concentração: Economia do Setor Público.

Aprovada em: ___ / ___ / _____

BANCA EXAMINADORA

Prof. Dr. Fabrício Carneiro Linhares (Orientador)
Universidade Federal do Ceará (UFC)

Prof. Dr. Ricardo Brito Soares
Universidade Federal do Ceará (UFC)

Prof. Dr. Vitor Hugo Miro Couto Silva
Universidade Federal do Ceará (UFC)

RESUMO

A ascensão da era digital e a proliferação de dispositivos conectados têm desencadeado uma avalanche de dados, conhecida como *big data*, que tem se revelado uma fonte promissora de evidências para a gestão pública. A pesquisa é motivada pelo crescente desafio de tentar desmistificar a utilização do *big data* por governos locais, onde existe uma baixa adesão do seu uso, visando aprimorar a eficiência das políticas públicas, a qualidade de vida de seus moradores e a sustentabilidade das cidades. Explorando as oportunidades e desafios inerentes à sua utilização. Neste contexto, considera-se imperativo examinar os diferentes aspectos envolvidos, desde a coleta e processamento dos dados até a formulação de políticas públicas embasadas em evidências, além de abordar os desafios tecnológicos, de equipe técnica qualificada, apoio político das lideranças locais, arcabouço legal, questões éticas de privacidade e segurança de dados, devem ser cuidadosamente consideradas ao implementar soluções baseadas em *big data*. Uma cartilha será apresentada para ilustrar como uma cidade poderá implementar uma solução de *big data*.

Palavras-chave: Big Data; Aprendizado de Máquina; Análise de Dados; Computação em Nuvem; Cidades Inteligentes.

ABSTRACT

The rise of the digital era and the proliferation of connected devices have triggered an avalanche of data, known as big data, which has proven to be a promising source of evidence for public management. The research is motivated by the growing challenge of demystifying the use of big data by local governments, where there is low adoption, aiming to enhance the efficiency of public policies, the quality of life for residents, and the sustainability of cities. Exploring the opportunities and challenges inherent in its use. In this context, it is considered imperative to examine the different aspects involved, from data collection and processing to the formulation of evidence-based public policies. Technological challenges, qualified technical staff, political support from local leaders, legal frameworks, ethical issues of privacy and data security must be carefully considered when implementing big data-based solutions. A guide will be presented to illustrate how a city can implement a big data solution.

Keywords: Big Data; Machine Learning; Data Analysis; Cloud Computing; Smart Cities.

SUMÁRIO

1	INTRODUÇÃO	9
2	ENTENDENDO O BIG DATA	13
2.1	Definição	13
2.2	Os V's do <i>big data</i>	13
2.3	Tecnologias essenciais para o <i>big data</i>	15
2.4	Tendências e padrões no uso de dados pela administração pública	18
3	DADOS: A MATÉRIA-PRIMA DO <i>BIG DATA</i>	21
3.1	Políticas de dados abertos	23
4	ANÁLISE DE DADOS – EXTRAINDO VALOR DOS DADOS	28
4.1	Aprendizado de máquina	30
4.1.1	<i>Processos de aprendizagem</i>	31
4.1.1.1	<i>Aprendizado supervisionado</i>	31
4.1.1.2	<i>Aprendizagem não supervisionada</i>	33
4.1.1.3	<i>Aprendizagem de reforço</i>	34
4.1.2	<i>Aprendizagem profunda (deep learning)</i>	35
4.2	Observações importantes a tomada de decisão baseada em dados	35
5	CIDADES INTELIGENTES	38
5.1	O caminho para cidades inteligentes	39
5.1.1	<i>Fatores essenciais na implantação de cidades inteligentes</i>	41
5.2	O uso de <i>big data</i> nas cidades inteligentes	42
5.3	Participação cidadã	44
6	DESAFIOS DO BIG DATA NA GESTÃO PÚBLICA	46
6.1	Apoio político	46
6.2	Quadro técnico limitado	47
6.3	Falta de integração e padronização dos dados disponíveis	48
6.4	Tecnológico e infraestrutura	50
6.5	Cibersegurança	51
6.6	Jurídicos e regulatórios	52
6.7	Considerações éticas relacionados à privacidade e proteção de dados pessoais	52
7	ETAPAS PARA A EXECUÇÃO DE UM PROJETO DE BIG DATA	54

7.1	Identificação de objetivos e desafios	54
7.2	Formar uma equipe qualificada	55
7.3	Avaliar os recursos existentes	55
7.4	Diagnóstico dos dados	56
7.5	Garantir a privacidade e a segurança dos dados	57
7.6	Preparação dos dados	57
7.7	Análise de dados	58
7.8	Desenvolva a aplicação	58
7.9	Apresentação e comunicação	59
7.10	Acompanhamento e aperfeiçoamento contínuo	59
7.11	Provedor em nuvem	59
8	ESTUDO DE CASO – BIGDATA FORTALEZA	65
8.1	Fortaleza Em Mapas	65
8.2	Mapa Colaborativo	67
8.3	BigData Fortaleza	67
9	CONCLUSÃO	71
	REFERÊNCIAS	74
	APÊNDICE – BIG DATA: OPORTUNIDADES E DESAFIOS PARA	
	GOVERNOS LOCAIS	77

1 INTRODUÇÃO

Poucas transformações trazem mudanças tão significativas para a sociedade como os avanços tecnológicos que ultrapassam as barreiras da inovação, modificando profundamente a forma de interação entre as organizações e às pessoas.

A medida que as cidades se expandem e suas populações aumentam, um planejamento urbano eficiente é essencial para lidar com os desafios desse crescimento, onde geralmente ocorre de forma desordenada, afetam as necessidades básicas que impactam a qualidade de vida de seus moradores.

Conforme as previsões da ONU¹, Organização das Nações Unidas, a alta densidade urbana deve continuar a aumentar, alcançando 70% da população, vivendo em cidades até 2050. No Brasil, segundo dados do IBGE², em 2022, as concentrações urbanas, que são arranjos populacionais ou municípios isolados com mais de 100 mil habitantes, abrigavam 124,1 milhões de pessoas, 61% da população.

De acordo com Bouskela *et al.*, (2016), uma das consequências da expansão dos espaços urbanos não é somente o fato de que as cidades são maiores, mas também que elas se confrontam com os problemas mais complexos, sendo imprescindível que haja maior preocupação em encontrar soluções para os problemas de infraestrutura urbana, para as questões sociais, culturais, ambientais, entre outros.

Atualmente, se está em um período impulsionado pelo crescimento das cidades e da evolução de tecnologias emergentes, como computação em nuvem, internet das coisas (IoT, sigla em inglês), armazenamento e processamento distribuídos.

A convergência desses dois fatores trouxe algo inovador, uma quantidade massiva de dados provenientes de diversas fontes, como mídias digitais, sensores, dispositivos IoT, dentre outros e que não para de crescer.

Projeções realizadas, afirmam que em abril de 2022, a internet atingiria 63% da população mundial, desse total, mais de 93% seriam usuários de redes sociais. A quantidade total de dados prevista globalmente em 2022 era de 97 zettabytes, um número projetado para crescer para 181 zettabytes até 2025³.

Como os grandes centros urbanos são os principais geradores de dados, surge então

¹ <https://news.un.org/pt/story/2019/02/1660701>

² <https://agenciadenoticias.ibge.gov.br/agencia-noticias/2012-agencia-de-noticias/noticias/37237-de-2010-a-2022-populacao-brasileira-cresce-6-5-e-chega-a-203-1-milhoes>

³ <https://www.domo.com/data-never-sleeps>

um novo mundo de oportunidades no enfrentamento dos desafios urbanos que pode ser encontrada por meio deles, que quando bem tratados e analisados, podem trazer evidências valiosas sobre a dinâmica das cidades e a qualidade de vida dos cidadãos, capacitando a gestão pública a estabelecer com maior grau de certeza e de forma mais embasada as suas decisões.

Diante desse contexto, as organizações públicas podem buscar soluções apropriadas para lidar com o crescente aumento do volume, da variedade e da velocidade dos dados em seu poder. Uma solução frequentemente associada a essa nova era, baseada em dados, é conhecida como *big data*, que transforma a forma como os governos operam e interagem com os cidadãos, melhorando a eficiência dos serviços públicos, a otimização de recursos, a antecipação de problemas e necessidades da população, além de proporcionar maior transparência na gestão pública.

Devido aos recursos limitados e as prioridades decorrentes dos desafios urbanos, muitas cidades tendem a demorar em adotar tendências tecnológicas como o *big data*, por isso, a importância que sejam desenvolvidos trabalhos que apresentem com clareza o potencial dessa tecnologia dentro das organizações públicas.

Com o advento da crise causada pela pandemia da COVID 19, foi posto em prova a relevância das tecnologias digitais como recurso imprescindível no auxílio as administrações públicas no combate as adversidades da pandemia. A procura de serviços públicos de forma *on-line* foi intensa pela população.

Diante disso, a Dissertação tenta responder a seguinte indagação: "Como os governos locais podem se beneficiar com o uso efetivo de *big data* para o desenvolvimento de cidades mais inteligentes?".

Segundo Davenport (2014), o *Big Data* desempenha um papel fundamental no processo de tomada de decisões, permitindo os gestores públicos a tomar decisões a partir de evidências valiosas encontradas na ampla variedade de fontes de dados em seu poder. Essas evidências podem ser usadas para melhorar a precisão de modelos preditivos e explanatórios, bem como para identificar tendências e padrões que podem ser usados para fundamentar decisões mais informadas.

As administrações públicas, possuem acesso a uma vasta quantidade de dados em suas bases de dados, trazendo uma oportunidade sem igual, que embasem os dados na tomada de decisão, desempenhando assim um papel essencial no aprimoramento de políticas públicas. São nas cidades onde os desafios da gestão pública estão mais próximos da população, que segundo dados do IBGE, em 2022, a maior parte da população do país (57% do total) habitava

apenas 319 municípios, o que, de acordo com a publicação, evidencia que as pessoas estão concentradas em centros urbanos acima de 100 mil habitantes.

O Ceará, por exemplo, possui 184 municípios e uma população residente de 8.791.688 pessoas, segundo censo de 2022, sendo que 56% da população vive em 176 municípios com até 100 mil habitantes.

Gráfico 1 - População por porte das cidades cearenses.



Fonte: IBGE, elaborado pelo autor

Porém, segundo indicadores do Mapa de Governo Digital⁴, somente 97 desses municípios possuem alguma estrutura organizacional para a área de TI, 149 deles têm conexão com a internet via cabo ou fibra ótica; apenas, 21 municípios possuem centro de controle de operações; e 10 possuem algum tipo de dispositivos IoT, como, semáforos inteligentes, sistemas de iluminação inteligente e sensores para monitoramento de risco.

Isso demonstra que apesar dos avanços, em determinado grau impostos pela crise sanitária, que ocasionou uma crescente demanda da população por serviços *on-line*, a quantidade de municípios cearenses, utilizados para essa análise, ainda é aquém do esperado.

Desse modo, o objetivo geral visa contribuir sobre a relevância do impacto na gestão pública com uso de tecnologias como o *big data* para o desenvolvimento de cidades mais inteligentes. A partir do objetivo geral foi definido os seguintes objetivos específicos:

- a) avaliar como a qualidade dos dados podem influenciar no processo de análise na hora da extração de evidências;

⁴ https://www.gov.br/governodigital/pt-br/transformacao-digital/rede-nacional-de-governo-digital/mapa/mapa_geral/

- b) sintetizar os métodos de análise de dados que podem ser utilizados na extração de evidências;
- c) examinar aspectos relacionados a privacidade e segurança no uso de *big data* na gestão pública;
- d) explorar como o uso de *big data* pode facilitar e promover a participação cidadã no processo de interação entre governo e sociedade;
- e) observar as oportunidades e desafios na transformação de cidades em ambientes mais inteligentes.

Além desta seção introdutória, contendo a contextualização geral do tema, este trabalho apresenta mais oito seções e um apêndice. A segunda seção inicia com uma abordagem teórica do conceito de *big data*. Na seção três foi realizada um estudo sobre as características e qualidades essenciais dos dados. Os procedimentos dos métodos de análises utilizados para transformar dado em valor foi descrito na quarta seção, seguida pela seção quinta que abordou os caminhos para criar uma cidade inteligente. Na sexta seção discutiu-se sobre os principais desafios para implantar o *big data* na gestão pública. Na sétima seção foi apresentado alguns passos que viabilizam um projeto de *big data*. Na oitava seção foram apresentados alguns casos de uso usados pela Prefeitura de Fortaleza e desenvolvidos pelo Instituto de Planejamento de Fortaleza. Na nona seção, as conclusões finais a respeito do estudo são apresentadas.

No apêndice deste trabalho, foi apresentada uma cartilha elaborada com o intuito de fornecer um guia prático e direcionado para gestores públicos. Esta cartilha se destina a oferecer uma compreensão clara e concisa sobre os benefícios do big data na gestão municipal, destacando sua relevância e potencial transformador. Usando uma forma acessível e objetiva, ilustrando como essa tecnologia pode ser aplicada para otimizar processos e promover o desenvolvimento urbano sustentável.

Ademais, o apêndice também incluiu um exemplo prático do uso da linguagem de programação Python em análises de dados. Esta demonstração servirá como um ponto de partida para gestores interessados em explorar as capacidades analíticas do big data utilizando ferramentas acessíveis e amplamente utilizadas. Assim, o apêndice não apenas complementarará o conteúdo teórico apresentado ao longo do trabalho, mas também oferecerá recursos práticos e tangíveis para aqueles que desejam iniciar sua jornada rumo à implementação eficaz do big data na gestão pública local.

2. ENTENDENDO O BIG DATA

2.1 Definição

No processo tecnológico se avalia que a crescente geração de dados nos dias atuais, vem sendo impulsionado pelo uso massivo dos meios digitais, tais como, redes sociais, blogs, sistemas web e sensores, que vem se tornando a principal fonte de matéria-prima para tomada de decisões nas instituições, sejam elas públicas ou privadas, mas junto com esse crescimento novos desafios apareceram.

Diversos autores apontam na literatura pesquisada, que a Internet representa um salto qualitativo de grande magnitude, transformando e redefinindo as formas de conhecer e relacionar-se do homem. O volume, a variedade e a velocidade cresceram tanto que os métodos tradicionais de processamento e armazenamento ficaram obsoletos, sendo necessário novas abordagens capazes de processar de maneira rápida e eficiente essa quantidade expressiva de dados da atualidade.

O papel e o impacto da tecnologia da informação e comunicação no mundo acadêmico moderno em geral, levou o conceito de *big data* a uma definição conjunta tão grande e complexa de dados que se torna um desafio para sistemas tradicionais de processá-los, armazená-los e analisá-los de forma eficaz e produtiva.

A Internet concebe uma nova forma de organizar a produção, o que na época com a revolução industrial, significava a fábrica como ponto de organizar a produção em massa.

Atualmente o *Big data* não é uma tecnologia específica, mas um conjunto de tecnologias em contínuo desenvolvimento ao longo do tempo, que agora são suportadas por infraestruturas capazes de lidar com a vasta quantidade e complexidade dos dados gerados em ambientes distribuídos e não convencionais, e a computação distribuída desempenhou um papel fundamental nesse processo.

Segundo Schönberger-Mayer e Cukier (2013, p.04), *big data* se refere a trabalhos em grande escala que não podem ser feitos em escala menor, para extrair novas ideias e criar novas formas de valor de maneiras que alterem os mercados, as organizações, a relação entre cidadãos e governos, etc.

2.2 Os Vs do *big data*

O termo "Vs do *big data*" refere-se a características-chave que são frequentemente usadas para descrever os desafios e as propriedades distintas do armazenamento, processamento e análise de grandes conjuntos de dados. Essas características não são estáticas e podem evoluir

ao longo do tempo em conformidade com os avanços tecnológicos.

Segundo o Gartner, em seu glossário define, "*Big data* são ativos de informações de alto volume, alta velocidade e/ou alta variedade que exigem formas inovadoras e econômicas de processamento de informações que permitem maior percepção, tomada de decisões e automação de processos".

A granularidade é uma característica importante que o *big data* nos dá, é o ponto de partida para compreender o que deverá ser extraído, com isso, a granularidade na medida certa, permite análises mais precisas. De acordo com Schönberger-Mayer e Cukier (2013, p.20), "Como se baseia em toda a informação, ou pelo menos no máximo possível, o *big data* permite que se analise detalhes ou se explore novas análises sem o risco sair do foco. Pode-se testar novas hipóteses em vários graus de granularidade."

A primeira característica é o **volume** caracterizado por sua grandeza, que excede a capacidade de armazenamento e processamento dos dados por meios tradicionais, referindo-se à quantidade de dados gerados a partir de diversas fontes. É uma característica relativa ao tempo e ao contexto que será utilizado, ou seja, o que representa grandeza hoje, amanhã pode não ser.

O segundo V é a **velocidade**, essa característica concentra-se na rapidez com que os dados são criados, coletados, transmitidos e necessitam ser processados para que decisões sejam tomadas em tempo real ou em um prazo muito curto de tempo.

Como o volume de dados pode ser gerado por diversas fontes e tipos, surgiu a **variedade** como a terceira característica. Devido as fontes de geração serem diversas, os dados não seguem padrões e nem tipos específicos, sendo que na atualidade os dados não estruturados são os principais tipos e as mídias sociais e os sensores são as principais fontes. Lidar com a análise dessa variedade de dados é uma das essências do *big data*, pois sistemas tradicionais não são adequados no processamento e armazenamento desses tipos de dados. Apesar dessa variedade ter suas complexidades, ela traz ao mesmo tempo explicações para determinados fenômenos que chegam a níveis tão profundos que a programação tradicional não seria capaz de encontrar.

A **veracidade** indica à confiabilidade dos dados. Como os dados são provenientes de diversas fontes, muitas delas difíceis de validar, eles podem conter erros, imprecisões ou até mesmo serem inverídicos, implicando em análises incorretas e tomada de decisões imprecisas. Essa característica está muito ligada a qualidade dos dados coletados. Em muitas análises de dados, garantir a veracidade das evidências encontradas, combinada com a velocidade que os dados são gerados e coletados representa um grande desafio para projetos de *big data*.

Não seria sensato todo o esforço em manter um grande volume, coletadas em diversas fontes de dados (variedade), com requisitos de velocidade em tempo real, mas que no final não seja capaz de extrair evidências que agregue valor para a organização. A entrega de **valor** é a característica que justificativa o projeto de *big data*, apesar de que alguns autores não a consideraram como característica, mas algo que é extraído no processo de análise de *big data*.

Segundo Vianna e Dutra (2016, p.193), "Valor: o benefício do investimento, o uso de *Big Data* exige um investimento que deve gerar retorno para as empresas como na melhoria da qualidade dos serviços e aumento da receita".

2.3 Tecnologias essenciais para o *big data*

Como dito anteriormente, *big data* não é uma tecnologia, mas a junção de diversas tecnologias que viabilizaram e impulsionaram o seu conceito. Juntas desempenham papéis fundamentais na capacidade para trabalhar com grandes volumes e variedade de dados, bem como a habilidade de processá-los em um tempo hábil, permitindo uma análise cada vez mais rápida e eficiente para a tomada de decisões. Não esquecendo que, o armazenamento e processamento distribuído, foram as chaves que viabilizaram o *big data*, e isso nos leva a um outro conceito fundamental, a escalabilidade.

Com o surgimento das ciências contemporâneas o uso das TIC's, novos e importantes fatos, como a visão de mundo tornou-se mais ampla, impulsionada pelas limitações geradas pela grande demanda de processamento e restrições físicas das arquiteturas tradicionais, a computação distribuída acenou como uma alternativa para amenizar os desafios computacionais tradicionais. Esse modelo detém atualmente um papel primordial no processamento, capaz de extrair informações relevantes das aplicações de *big data*.

O volume sendo uma característica relativa, podendo variar de acordo com o contexto do negócio ou no tempo, exige cargas de processamento e armazenamento conforme a demanda da aplicação. Para não prejudicar o desempenho, a escalabilidade horizontal (*scale-out*), fruto da computação distribuída, foi a solução encontrada, onde a capacidade de um sistema é aumentada com a adição de mais recursos para atender as necessidades do negócio, diferente da escalabilidade vertical (*scale-up*) que prioriza o aumento do desempenho de um recurso, como exemplo, adição de mais processador, memória ou armazenamento de uma determinada máquina.

A escalabilidade horizontal permite um processamento paralelo da carga de trabalho, onde uma tarefa pode ser executada simultaneamente em diferentes máquinas, além de que são

mais tolerantes a falhas, pois a falha de uma máquina não paralisa o sistema.

Esse processamento distribuído é realizado em clusters que nada mais são do que um conjunto de computadores interconectados, que quando combinados, criam uma única entidade de processamento de alto desempenho, cada computador do cluster é chamado de “nó”. Os clusters são usados para melhorar o desempenho, a confiabilidade e a escalabilidade de aplicações, que exigem um grande poder de processamento, como análise de dados em *big data*.

Embora a computação distribuída seja algo promissor, algumas de suas características dificultam sua utilização, pois dividir uma tarefa em subtarefas e então executá-las paralelamente em diversos nós de processamento, não é uma tarefa trivial. Foi nesse contexto que foi desenvolvido o *Apache Hadoop*, um *framework* para o processamento e armazenamento de grandes quantidades de dados em clusters de forma ágil, escalável e tolerante a falhas, sendo capazes de manter a integridade dos dados e tudo isso de maneira transparente para o usuário.

Desenvolvido como um projeto de código aberto, a principal motivação por trás do *Hadoop* era lidar com o crescente volume de dados na era da internet, oferecendo uma solução escalável e confiável para o processamento de *big data*, permitindo a execução de operações em paralelo em um cluster de máquinas (processamento distribuído).

Esse poder de processamento é fundamental para a eficiência do desenvolvimento e treinamento de modelos de aprendizado de máquina que utilizam algoritmos que executam cálculos complexos em grandes conjuntos de dados em tempo hábil para a tomada de decisões.

Ademais, o armazenamento tradicional, que trata de dados estruturados, geralmente usados pelos bancos de dados relacionais, não são mais capazes de absorver dados gerados por fontes como redes sociais e sensores, por exemplo, que em sua grande maioria criam dados não estruturados ou semiestruturados. Para solucionar esse problema surgiu como um paradigma não-tradicional os bancos de dados *NoSQL (Not Only SQL)*, capazes de lidar com os grandes volumes e variedade de dados e resolver alguns dos desafios que surgiram para implementar projetos de *big data*.

A importância do “Not Only” SQL explica-se pelo modelo relacional ser baseado no fato de que, na época de sua criação, início dos anos 70, acessar, categorizar e normalizar dados era bem mais fácil que hoje. Praticamente não existiam dados não estruturados circulando pelos computadores da época. Também não foi desenhado para escala massiva nem processamento extremamente rápido. Seu objetivo básico era possibilitar a criação de queries que acessassem bases de dados corporativas e, portanto, estruturadas. (Taurion, 2013, p. 21).

A Tecnologia da Informação e Comunicação (TIC's) tem contribuído imensamente para melhorias sociais e econômicas, tais como maior emprego e produtividade, aumentando o acesso a uma maior qualidade de vida.

Bancos de dados *NoSQL* não exigem um esquema de dados rígido, o que os torna mais flexíveis para armazenar e gerenciar dados de diferentes formatos e estruturas, além de ser desenhado para aumentar a sua escala em sentido horizontal (armazenamento distribuído), por meio de clusters, essa adaptabilidade, tem tornando os bancos de dados *NoSQL* em uma solução, para tratar dados não estruturados, superando os problemas relacionados ao armazenamento de volumes e variedades de dados cada vez maiores nos dias atuais.

Uma solução alternativa em adquirir e manter infraestrutura de TIC localmente (os chamados ambientes *on-premise*) é a contratação de recursos computacionais como processamento, armazenamento, dentre outros, disponibilizados pela internet por meio de provedores de nuvem. A computação em nuvem fornece o poder de escalar recursos e serviços, permitindo que as organizações contratem sob demanda, tornando-se uma escolha mais eficaz e econômica, fornecendo armazenamento e/ou processamento distribuído.

Pode-se pensar que a computação em nuvem é também um impulsionador para Big Data, pois podem-se usar nuvens públicas para suportar imensos volumes de dados e as características de elasticidade das nuvens permitem que acionemos servidores virtuais sob demanda, apenas no momento de tratar estes dados." (Taurion, 2013, p. 71)

As TIC's incorporam tecnologias e técnicas eletrônicas usadas para gerenciar informações e conhecimento, incluindo ferramentas de manuseio de informações usadas para produzir, armazenar, processar, distribuir e trocar informações, para tanto o volume de dados utilizados nas análises em projetos de *big data* pode variar em determinado momento, onde picos de processamento e armazenamento podem ocorrer. A capacidade de escalar recursos de acordo com a demanda desse volume é um aspecto fundamental para acomodar com sucesso o atual ambiente de negócio.

Os benefícios das TIC's podem ser alcançados por intermédio da melhoria da prestação de serviços de saúde e prevenção de doenças, ou por meio de melhor infra-estrutura social, crescimento econômico ou outros fatores mais abrangentes da saúde da população.

Com computação em nuvem se tem a flexibilidade de escalar a capacidade computacional de acordo com as necessidades do momento sendo uma solução muito mais econômica e eficiente do que manter uma infraestrutura local.

Os provedores de soluções em nuvem oferecem desde servidores virtuais até

serviços *servless* de análise e aprendizado de máquina, permitindo assim que as organizações façam escolhas mais adequadas conforme suas necessidades.

Existem três modelos básicos de entrega de recursos e serviços, são eles:

- a) infraestrutura como serviço (IaaS): é o modelo que mais se aproxima de uma infraestrutura local. Seria como ter sua própria infraestrutura virtualizada, podendo incluir servidores virtuais, armazenamento, redes e outros componentes de infraestrutura;
- b) plataforma como serviço (PaaS): esse modelo oferece um ambiente completo para o desenvolvimento de aplicações na nuvem. Os desenvolvedores conseguem criar, gerenciar e hospedar suas aplicações sem se preocupar com atualizações de softwares usados para o desenvolvimento e nem a infraestrutura necessária para isso;
- c) software como serviço (SaaS): neste modelo, um *software* é entregue pela internet como um serviço. Os clientes o acessam através da internet, sem a necessidade de ter uma licença para uso, seria como alugar um software pela internet.

A internet das coisas (IoT) é a capacidade de conectar à internet, sensores usados no dia a dia das pessoas e organizações, onde se pode coletar e transmitir dados, criando um ambiente mais inteligente. Com a crescente adoção de dispositivos IoT a quantidade de dados coletados em tempo real aumentou significativamente, tornando-se um aspecto crucial para análises de *big data*, permitindo a detecção de padrões, tendências e *insights* instantâneos.

Para Taurion (2013, p. 30), “A crescente miniaturização da tecnologia bem como o aumento da sua capacidade de processamento e armazenamento permite a criação da Internet das Coisas, o que aumentará de forma exponencial a geração de dados.”

Com o avanço da quantidade de dispositivos conectados à internet, surge a necessidade de um aprimoramento das redes móveis, a tecnologia **5G** é a próxima geração de redes de comunicação móvel que oferece velocidades muito mais rápidas, com menor latência e maior capacidade de conexão em comparação com as gerações anteriores, sendo fundamental para lidar com o volume de dados gerados pelos dispositivos IoT.

Existem várias infraestruturas de conectividade (fibra ótica, 3G e 4G, exemplos) que viabilizam vários projetos de big data, mas a adoção combinada das tecnologias de IoT e 5G são requisitos essenciais na infraestrutura de cidades inteligentes.

2.4 Tendências e padrões no uso de dados pela administração pública

Estimulada pela disponibilidade de dados e avanços tecnológicos algumas tendências e padrões refletem o crescente reconhecimento do valor do *big data*, fazendo com que a administração pública possa ter maior precisão e agilidade nas suas tomadas de decisões, aumentando sua expectativa e satisfação perante a população.

Dessa forma, se esclarece que a inovação tecnológica, significa chegar a novas ideias, sobre como fazer as coisas melhor ou mais rápido. Trata-se de fazer um produto ou oferecer um serviço que ninguém tinha pensado antes. É sobre colocar novas ideias para trabalhar na empresa e ter uma força de trabalho qualificada, que possa usar essas ideologias.

Como exemplo cita-se a administração pública baseadas em dados, tomam suas decisões com mais evidências e com menos “achismo”. Como os dados são fatos concretos, a prática em utilizá-los, ajuda a reduzir decisões que poderão sofrer alguma influência de viés pessoal, intuição ou opiniões subjetivas. Um padrão comum reconhecido nas administrações modernas é o uso rotineiro de algoritmos de aprendizado de máquina na realização de suas análises, que permite concretizar previsões mais assertivas e *insights* mais profundos a partir dos grandes conjuntos de dados, representando uma evolução das possibilidades para criação e avaliação de políticas públicas.

O uso de dados pela administração pública permite avançar em determinadas áreas da própria gestão, melhorando a colaboração interna entre setores, por meio do compartilhamento de dados, otimizando os processos internos, melhorando a eficiência operacional, capacitando o monitoramento em tempo real por meio da coleta e análise de dados, permitindo assim, respostas mais rápidas e eficientes em situações como desastres naturais, surtos de doenças, por exemplo.

Além disso, aumenta a transparência governamental, como por exemplo, criando plataformas de dados abertos, onde permite a população acessar e entender os dados gerados pela administração pública, promovendo assim uma participação cidadã mais ativa no processo da gestão. Mas à medida que mais dados são coletados e disponibilizados, a privacidade e a segurança se tornam uma preocupação crescente, sendo necessário que sejam estabelecidas regras e políticas visando a proteção dos dados sensíveis da população.

Outra tendência fundamental para aproveitar ao máximo os recursos e conhecimentos disponíveis em projetos dessa natureza é a criação de equipes multidisciplinares, ou ainda a formação de parcerias com empresas privadas ou academia pelas organizações públicas. E sempre tem o cidadão como principal beneficiário, mesmo que indiretamente, pois o uso de *big data* não deve ser encarado como a utilização de tecnologia por tecnologia, mas

sim uma forma de melhorar o acesso e a prestação de serviços públicos a população.

Com o conhecimento e as ferramentas corretas, o servidores poderão modificar o núcleo central de sua atividade diária, de seu desenvolvimento, de seu relacionamento com os usuários e até, e mais importante, com os diferentes operadores legais com os quais lida todos os dias, de tal maneira que possa avançar exponencialmente os procedimentos obrigatórios.

3. DADOS: A MATÉRIA-PRIMA DO *BIG DATA*

A Tecnologia contemporânea leva o ambiente administrativo para uma evolução em que a profissionalização e a especialização são necessárias, não somente para os setores públicos, mas para todos os usuários.

Em maio de 2017 a revista *The Economist* publicou que o recurso mais valioso do mundo já não era mais o petróleo, mas os dados. Pode-se conceituar dados como valores atribuídos a algo que ainda não foi processado, ou seja, assim como o petróleo que tem que passar por um processo de refino para produzir o seu verdadeiro valor os dados em estado bruto, também precisam passar por um processo de tratamento, onde geraram informações que serviram como insumos para tomada de decisão mais assertivas.

Apesar de ser a matéria-prima para tomada de decisões, é necessário extrair valor desses dados. O uso desse valor extraído embasa os gestores públicos nas tomadas de decisões tanto operacionais como estratégicas da organização, além de obter informações sobre as expectativas e necessidades da sociedade.

Ilustra-se que toda inovação tecnológica abrange uma estrutura legal dentro do qual atuar, a proteção dos negócios, o mecanismo corporativo, a proteção fiscal, trabalhista, de defesa do consumidor e o registro de ativos intelectuais, como direitos autorais, propriedade industrial, tais como, marcas e patentes, registro de nomes de domínio, proteção de dados pessoais, sigilo industrial, confidencialidade, exclusividade, formas modernas de contratação internacional, entre outros aspectos legais.

As pesquisas de Kim, Trimi e Chung (2014) apresentam que, com um vasto volume de dados armazenados em seus bancos de dados, o governo possui a oportunidade de analisá-los minuciosamente e extrair *insights* que podem resultar na criação de serviços inovadores e no aprimoramento de sua administração.

A variedade do *big data* trouxe a possibilidade de coletar dados de várias fontes, que podem ser gerados em formatos diferentes, que definem como poderão ser armazenados, acessados e processados. Dispositivos de IoT, *smartphones*, redes sociais, registro de atividades em *websites* são só alguns exemplos dessas fontes de dados.

Quando os dados são organizados em uma estrutura rígida, padronizada e organizada geralmente em formato de tabelas ou campos, de modo que cada dado tem um tipo pré-definido e se encaixa em uma categoria específica, são os chamados dados estruturados. São muitos utilizados em bancos de dados tradicionais, como os relacionais.

Muitas das fontes de dados ditas “modernas” são caracterizadas por sua diversidade e falta de uniformidade em termos de formato e conteúdo, gerando tipos de dados que não seguem uma estrutura rígida. Como não possuem um esquema fixo, os dados não estruturados podem assumir várias formas, como texto, imagens, vídeos, áudios, e etc.

Sendo o principal tipo gerado na atualidade, era um problema para os bancos de dados tradicionais que só foram solucionados com o surgimento de bancos *NoSQL*, mas que ainda representam um desafio na hora da análise desses dados, pois em determinadas ocasiões é necessário a utilização de técnicas avançadas de aprendizado de máquina para extração de valor, como no caso de análise de sentimento em dados gerados por redes sociais, fóruns *on-line* ou blogs.

Existe um tipo de dados que não seguem uma estrutura rígida, como os dados estruturados, mas possuem alguma organização, representando um meio-termo entre os outros dois tipos. São chamados de dados semiestruturados, encontrados geralmente em fontes de dados em que a flexibilidade é necessária, mas algum grau de organização ainda é desejado, geralmente sensores usam esse tipo de estrutura.

Como atualmente a maioria dos dados gerados são do tipo não estruturado, que possuem como característica a heterogeneidade de formatos, a simples coleta não é mais o bastante. As organizações devem ter a preocupação com a qualidade desses dados coletados. O valor como uma das características-chave do *big data* está ligado diretamente a qualidade dos dados e para que eles atendam a esse requisito algumas características são necessárias:

- a) precisão: o dado deve ser exato, confiável, devendo ser livres de erros;
- b) disponibilidade no tempo: devem refletir a realidade do momento, para não levar a decisões baseadas em dados obsoletas;
- c) relevância: os dados devem refletir a realidade do contexto em que serão usados;
- d) integralidade: os dados não devem ser corrompidos ou incompletos de forma a prejudicar a sua utilidade;
- e) simplicidade: os dados devem ser apresentados da forma mais clara e compreensível;
- f) confiabilidade: os dados não devem mudar de forma inesperada ou inconsistente.

Antes de extrair valor dos dados, algumas etapas devem ser citadas, a parte da coleta, onde os dados brutos são concentrados, depois vem a etapa mais demorada e fundamental que seria o pré-processamento onde a preparação, o tratamento e a limpeza devem ser conduzidas, a etapa seguinte é da análise propriamente dita, momento em que vários padrões

são identificados e por último a visualização dos resultados, que deve ser conduzida de uma maneira fácil de visualizar e interpretar. Na ausência de uma supervisão nessas etapas a qualidade dos dados pode ser comprometido, assim como análises imprecisas ou tendenciosas poderão ser apresentadas.

Para garantir que isso não ocorra, é necessário a criação de políticas e procedimentos em cada etapa, a fim de evitar impactos negativos. Então, não basta ter uma excelente capacidade para coleta, armazenamento e análise de dados se a qualidade for comprometida. Pensando em mitigar esses impactos negativos, criou-se um conjunto de padrões e políticas chamada de governança de dados, que possui como princípio o gerenciamento do ciclo de vida dos dados, desde a coleta até o descarte, pois dados de qualidade representam o principal insumo em projetos de *big data*.

A governança de dados é uma estratégia essencial em organizações que trabalham com dados, basicamente deve coordenar todo o ciclo de vida dos dados, desde a coleta até o descarte, garantindo que os mesmos estejam seguros, disponíveis, mas também privados.

Mas para que usar esses dados? Só a gestão pública poderá responder essa pergunta, mas políticas orientadas a dados fortalecem e melhoram a tomada de decisões, auxiliando na mudança de uma postura reativa para uma postura proativa, sendo capaz de entregar a população serviços públicos de forma simples, de fácil acesso, inteligente e personalizado.

3.1 Política de dados abertos

A política de dados abertos visa garantir e facilitar à sociedade o acesso aos dados produzidas ou custodiadas pelos órgãos públicos, com o intuito de promover a transparência, a participação popular e até mesmo o desenvolvimento de serviços ou aplicações pelo uso ou reuso desses dados.

Gomes e Fertig (2016) relatam que a política de livre acesso a dados públicos tem como objetivo, permitir que qualquer indivíduo ou entidade tenha a capacidade de acessar e empregar informações produzidas ou coletadas pelo governo. Essa abordagem tem o potencial de acarretar uma série de vantagens para a cidade e a sociedade, impulsionando o progresso econômico e aumentando a consciência e a segurança dos cidadãos.

Segundo a *Open Knowledge International*, “Dados abertos são dados que podem ser livremente usados, reutilizados e redistribuídos por qualquer pessoa – sujeitos, no máximo, à exigência de atribuição da fonte e compartilhamento pelas mesmas regras.”

Para o Tribunal de Contas da União (2015, p.7), os princípios fundamentais para os

dados abertos governamentais, são:

- a) completos: todos os dados públicos estão disponíveis, ou seja, dado público é aquele que não está sujeito a limitações válidas de privacidade, segurança ou controle de acesso;
- b) primários: os dados são apresentados tais como os coletados na fonte, com o maior nível de granularidade e sem agregação ou modificação;
- c) atuais: são disponibilizados tão rapidamente quanto necessária à preservação do seu valor;
- d) acessíveis: são disponibilizados para o maior alcance possível de usuários e para o maior conjunto possível de finalidades;
- e) compreensíveis por máquinas: são razoavelmente estruturados de modo a possibilitar processamento automatizado;
- f) não discriminatórios: são disponíveis para todos, sem exigência de requerimento ou cadastro;
- g) não proprietários: são disponíveis em formato sobre o qual nenhuma entidade detenha controle exclusivo;
- h) livres de licenças: não estão sujeitos a nenhuma restrição de direito autoral, patente, propriedade intelectual ou segredo industrial, mas as restrições sensatas relacionadas à privacidade, segurança e privilégios de acesso devem ser permitidas

Apesar de apresentar características relacionados, contribuindo para o aumento da transparência e o acesso às informações governamentais, portais da transparência e políticas de dados abertos possuem focos diferentes.

O dinamismo das transformações, o aprimoramento da tecnologia e a agilidade da globalização nos meios de comunicação, requerem de qualquer profissional, inclusive os que atuam no cenário governamental, têm um perfil aberto a novas ideias e conceitos e de valorização aos conhecimentos de cada componente da equipe para contar com a participação de todos e estabelecer uma gestão em excelência, desprendida e emancipada.

Com a inteligência da máquina vários aspectos da lei e da tecnologia da informação, acelerou uma maior transparência, como os Portais da transparência, que têm como foco disponibilizar informações sobre as atividades financeiras e administrativas do governo, isso inclui, por exemplo, contratos, licitações, salários de servidores, permitindo que a sociedade faça o acompanhamento desses gastos. Já as políticas de dados abertos têm um foco

mais amplo, visando promover a abertura de dados governamentais de maneira geral, não se limitando apenas com informações financeiras e administrativas.

A abertura de dados públicos não aumenta apenas a transparência, mas também a participação popular, ao promover a circulação dos dados entre as organizações públicas e a sociedade, cria-se assim um ambiente de reuso desses dados. Permitir que os dados sejam reutilizáveis cria um ambiente onde novos produtos ou serviços possam ser desenvolvidos pela sociedade, impulsionando assim um ciclo de inovação. Dessa forma as organizações públicas podem se beneficiar com os dados coproduzidos pela sociedade para elaboração de políticas ou serviços públicos.

Quando essa ampla variedade de dados do governo é aberta, fomenta a participação popular. Segundo o Tribunal de Contas da União (2015, p.13), assim organizações, cidadãos, acadêmicos e até mesmo instituições públicas têm a possibilidade de utilizar bases de dados públicos para a produção e o compartilhamento de novos conhecimentos e de novos serviços, numa concepção de coparticipação entre ente privado e governo na oferta de serviços públicos à sociedade.

Para Zuiderwijk e Janssen (2014), as políticas voltadas para a abertura de dados têm o potencial de ampliar a participação e promover a inclusão social, ao mesmo tempo em que incentivam o uso mais amplo dos dados com a implementação da política.

Desse modo, uma questão que deve ser observada nas políticas de dados abertos, é a garantia para o acesso aos dados, que seja amplo, incluindo também, grupos desfavorecidos ou grupos que possam ter dificuldades no acesso à tecnologia, sendo necessário a inclusão social e digital desses grupos por parte do setor público. No entanto, esse potencial só será totalmente explorado se os dados públicos forem verdadeiramente abertos, sem restrições, legais, financeiras ou tecnológicas, para garantir a inclusão de todos na reutilização desses dados

Os governos que utilizam dados abertos devem estabelecer políticas claras de como esses dados se delineiam, os princípios, objetivos e diretrizes para a disponibilização de dados, criando um compromisso oficial e transparente, e sempre se certificar de que os dados publicados não incluam informações pessoais ou sensíveis que possam violar a privacidade dos cidadãos.

Essa transformação não só dificulta a proteção da privacidade como também apresenta uma ameaça totalmente nova: castigos com base em propensões, isto é, a possibilidade de usar previsões de big data sobre pessoas para julgá-las e puni-las antes mesmo que elas ajam, o que renega a ideia de justiça e livre-arbítrio. Além da privacidade e da propensão, há um terceiro perigo. Corremos o risco de sermos vítimas da ditadura dos

dados, na qual adoramos as informações e os resultados de nossas análises e acabamos usando-os de forma equivocada. Com responsabilidade, o big data é um instrumento útil de tomada de decisão. Se usados sem sabedoria, eles se tornam um instrumento de poderosos, que podem transformá-los numa fonte de repressão, seja ao frustrar clientes e funcionários ou, pior, ao atacar cidadãos.” Schönberger-Mayer e Cukier (2013, p. 105 - 106).

Uma forma de garantir a privacidade de dados pessoais é o uso de técnicas de anonimização e pseudonimização, ambas previstas na Lei Geral de Proteção de Dados (LGPD), segundo a lei “anonimização: utilização de meios técnicos razoáveis e disponíveis no momento do tratamento, por meio dos quais um dado perde a possibilidade de associação, direta ou indireta, a um indivíduo;” e “dado anonimizado: dado relativo a titular que não possa ser identificado, considerando a utilização de meios técnicos razoáveis e disponíveis na ocasião de seu tratamento;”.

Ainda conforme a lei “... a pseudonimização é o tratamento por meio do qual um dado perde a possibilidade de associação, direta ou indireta, a um indivíduo, senão pelo uso de informação adicional mantida separadamente pelo controlador em ambiente controlado e seguro.”.

Para se compreender, se esclarece que em 28 de dezembro de 2018, a Medida Provisória no. 869/2018, que alterou determinadas disposições da LGPD e criou a Autoridade Nacional de Proteção de Dados (ANPD), entre outras modificações, a LGPD entrou em vigor em agosto de 2020, em vez de fevereiro de 2020, conforme necessário, quando a LGPD foi publicada pela primeira vez

A utilização de dados abertos capacita a administração pública na utilização do *big data*, ao promover credibilidade regulatória do acesso e uso. Como representa mudanças significativas no ambiente organizacional da gestão e na própria relação entre governo e sociedade, para que projetos baseados em dados tenham sucesso, é necessário que seja criada com antecedência uma cultura governamental capaz de assimilar esse novo paradigma.

Tabela 01 – Fontes de dados abertos

Portais de dados abertos	
Portal Brasileiro de Dados Abertos	https://dados.gov.br/home dados publicados pelo governo federal e por governos locais para realizar pesquisas, desenvolver aplicativos e criar novos serviços.
Sistema IBGE de Recuperação Automática	https://sidra.ibge.gov.br/home/ipp/brasil permite a consulta aos dados armazenados no Banco de Tabelas

(SIDRA)	Estatísticas.
Banco Interamericano de Desenvolvimento (BID)	https://data.iadb.org/DataCatalog/Dataset indicadores de projetos realizados pelo banco em alguns países.
Centro Regional de Estudos Desenvolvimento da Sociedade da Informação	https://data.cetic.br/ uso das tecnologias de informação e comunicação (TIC) no Brasil.
Banco Central	https://dadosabertos.bcb.gov.br/ o meio utilizado pelo BC para disponibilizar dados e informações públicas.
Ipea	http://www.ipeadata.gov.br/Default.aspx dados e indicadores sobre distribuição de renda, pobreza, educação, saúde, previdência social e segurança pública.
Portal da Transparência	https://portaldatransparencia.gov.br/download-de-dados aqui é possível baixar os dados apresentados no Portal da Transparência do Governo Federal.
Dados Abertos Fortaleza	https://dados.fortaleza.ce.gov.br/ informações sobre as mais diversas áreas da cidade.

Fonte: Elaborada pelo auto

A verdadeira essência não se encontra nos dados e sim na maneira como eles serão usados, ao alterar o volume, modifica a essência da análise, noções de causalidade abrem caminho para a correlação, onde mais pistas são examinadas sem que precise ter um entendimento aprofundado dos fatos.

4. ANÁLISE DE DADOS – EXTRAINDO VALOR DOS DADOS

Sinaliza-se que os governos possuem um potencial de acumular uma grande quantidade de dados de suas inúmeras operações, mas a acumulação desses dados por si só não é suficiente para promover melhoria na tomada de decisão. A capacidade de extrair conhecimento desses dados é onde reside o verdadeiro valor na tomada de decisões informadas, e isso é realizado através de um processo chamado de análise de dados.

A expressão “Analítico” é um termo que se relaciona com o ato de examinar, estudar ou investigar algo em detalhes e a partir daí conseguir extrair algo significativo. Na análise de dados, o enfoque analítico refere-se à capacidade de examinar conjuntos de dados de forma aprofundada para identificar padrões e tendências, habilitando a extração de *insights* úteis para tomadas de decisões mais assertivas.

Essa não é uma prática nova, em décadas passadas já era utilizado dados no processo de tomada de decisão, mas pela escassez de dados e da ausência de tecnologias mais avançadas esse processo geralmente combinava fatores de cunho mais pessoal, envolvendo opiniões e sugestões de grupos de interesse ou políticos, além do conhecimento de alguns servidores com base em experiências passadas.

Antes do *big data*, a análise geralmente se limitava a uma pequena quantidade de hipóteses que se definia bem antes de se coletar os dados. Quando se deixa que os dados falem por si, pode-se gerar conexões que nem se sabia que existiam (Schönberger-Mayer e Cukier, 2013, p.09)

Uma forma de impulsionar o uso de *big data* na gestão pública é revelar o valor dos dados, descobrindo evidências. Entretanto, antes é necessário ter uma compreensão clara do que poderá ser alcançado ao utilizar dados, devem ser feitas perguntas objetivas sobre quais decisões, políticas públicas ou programa, se desejar melhorar ou fortalecer com a análise de dados.

A capacidade de fazer perguntas relevantes para definir quais problemas serão resolvidos é o passo fundamental para determinar quais ferramentas, técnicas e modelos preditivos deveram ser utilizados de forma adequada para um processo bem-sucedido de análise de dados.

Quando examinada, a área de análise de dados poderá ser dividida em quatro tipos de abordagens diferentes, duas delas revelam o que está acontecendo agora e o que aconteceu no passado. Vale lembrar que a divisão da análise de dados em categorias não é rígida e pode haver sobreposição entre elas.

A primeira delas é a **análise descritiva**⁵(o que aconteceu?). É o tipo mais antigo e comum usado, consiste em revelar o que os dados dizem, utilizando dados históricos que foram coletados, relatando assim o que aconteceu e o que está acontecendo. As principais maneiras de descrever os dados são por meio da utilização de medidas de tendência central (como exemplos, média e mediana) e de medidas de dispersão (como exemplos, variância e desvio padrão), que fornecem uma representação clara e compreensível dos dados.

A **análise diagnóstica** (por que isso aconteceu?) enfatiza a descoberta de padrões e tendências, tentando entender as causas principais e correlacioná-las, envolvendo a identificação de relações de causa e efeito entre variáveis. São úteis para deduzir e inferir o sucesso ou o fracasso de iniciativas centradas em dados, compreendendo as razões dos padrões e tendências identificados na análise descritiva.

Já as outras duas, realizam uma projeção de um resultado futuro a partir da investigação de dados passados, o sucesso dessas análises se deve ao *big data*, pois quanto mais dados forem oferecidas para essas análises mais precisas serão as previsões. Como elas são baseadas em probabilidade nunca serão completamente precisas, ou seja, elas preveem o que pode acontecer no futuro com um nível aceitável de confiabilidade que inclui cenários hipotéticos e avaliação de riscos.

Na **análise preditiva** (o que vai acontecer?), embora seja baseado em dados históricos, envolve a construção de modelos e técnicas de aprendizado de máquina para realizar previsões de um evento futuro ou uma tendência, a partir da identificação de padrões passados em sua base de dados, esse tipo de análise permite o mapeamento de possíveis acontecimentos futuros.

Já na **análise prescritiva** (o que se pode fazer caso isso aconteça?) resume em entender o que é preciso fazer dada uma previsão futura, ela não apenas prevê o que pode acontecer, mas também sugerem o que deve ser feito, traçando as possíveis consequências de cada ação. É a análise mais sofisticada de todas, utilizando uma abordagem baseada em dados que utiliza técnicas avançadas de modelagem, otimização e simulação para recomendar ações específicas que levem a resultados desejados.

Todo o processo para se chegar a esse ponto de diversos cenários e *insights* está diretamente relacionado com a qualidade e volume de dados que se tem para entender o passado. A construção de um modelo analítico, a escolha de um algoritmo de aprendizagem de máquina

⁵ https://online.hbs.edu/Documents/a-beginners-guide-to-data-and-analytics.pdf?_gl=1*d25jll*_gcl_au*MjA0MDkxOTUxMC4xNzA4NDI1NzYy&_ga=2.194118317.1085675346.1708425763-1663992936.1708425762

e uma compreensão sólida das nuances do negócio, também são importantes nesse processo, pois caso não sejam bem conduzidos, corre-se o risco de decisões erradas, que podem levar a rumos perigosos na tomada de decisões.

Entretanto, de toda forma deve ser garantido que essas análises sejam implementadas de forma responsável, pois o uso dos dados para suas previsões deve levar sempre em consideração as questões éticas e de privacidade de como esses dados serão usados, resguardando à legalidade que abrangem a ética e a privacidade pessoal.

4.1 Aprendizado de máquina

O aprendizado de máquina (*machine learning*, em inglês) é um método de análise de dados que automatiza a construção de modelos analíticos, utilizando algoritmos que interativamente aprendem a partir de dados, permitindo descobertas de *insights*, que podem levar a novas descobertas não evidentes em análise tradicional. É um subconjunto da inteligência artificial que tenta aprender por intermédio do treinamento de modelos usando grandes conjuntos de dados (*big data*) para que possam aprender a reconhecer e generalizar padrões, com o objetivo de tomar decisões ou fazer previsões precisas ao utilizar novos dados.

Segundo Taurion (2013, p. 35), “Uma “*learning machine*” aprende com dados e quanto mais dados, mais o algoritmo aprende. Cria-se, portanto, um círculo virtuoso. *Big Data* é um passo significativo em busca da computação cognitiva.”

A diferença fundamental entre sistemas convencionais e sistemas baseados em aprendizado de máquina está na forma como eles são projetados para resolver problemas ou realizar tarefas. Sistemas convencionais são programados de maneira explícita, com algoritmos e regras definidos pelos desenvolvedores, se concentra no código, por outro lado, os sistemas baseados em aprendizado de máquina são projetados para aprender e melhorar a partir dos dados para aprimorar seu conhecimento sobre um tópico específico, se concentra nos dados.

Para programar algoritmos de aprendizado de máquinas é necessário um conjunto de dados grande o suficiente para que padrões possam ser reconhecidos e então aplicar técnicas estatísticas e robusta nesse grande conjunto de dados.

Mas para a sua eficácia, o aprendizado de máquina depende muito de dados para treinar seus modelos e realizar previsões, sendo necessário que os dados usados tenham qualidade, sejam confiáveis e acessíveis. Previsões imprecisas podem surgir caso esses dados sejam de baixa qualidade, pois faria que o algoritmo aprendesse com padrões imprecisos que levariam a resultados incorretos e em alguns casos previsões discriminatórias. A performance

de uma análise de dados eficiente dependerá sempre do fornecimento de dados de boa qualidade.

Os algoritmos de aprendizado de máquinas não são novidades, eles existem a algumas décadas, mas a grande mola propulsora de toda essa mudança foi o *big data*, a aprendizagem só se tornou eficaz devido ao grande volume de dados, da evolução das técnicas analíticas, da análise de dados em tempo real e do desenvolvimento de aplicações inteligentes.

Esses algoritmos quando reconhecem padrões nos dados são capazes de auxiliar os gestores públicos nas suas tomadas de decisões melhorando assim a eficiência geral dos serviços ofertados a população, a análise preditiva pode ser usada para prever tendências que direcionam os governantes a se adiantar na demanda por serviços em casos de possíveis picos ou na prevenção de surtos sanitários.

Big data não é apenas um grande conjunto de dados, mas também uma grande variedade, e essa variedade traz complexidade e ao mesmo tempo explicações para determinados fenômenos a níveis tão profundos que a programação tradicional não seria capaz de processar, além da qualidade encontrada nessa quantidade, que são as peculiaridades escondidas e preservadas nessas fontes de dados.

Como cada algoritmo de aprendizado de máquina possui características e propriedades diferentes a escolha certa dependerá do tipo de problema que se tenta resolver. O tipo de problema a ser resolvido, determina o tipo de algoritmo a ser utilizado, o uso certo do algoritmo para manipular dados é a pedra fundamental do aprendizado de máquinas. Então, a identificação com o maior grau de precisão possível do problema é fundamental, pois quanto mais precisa a pergunta sobre o problema for, mais precisa será a resposta e, portanto, maior será o valor atribuída a ela.

4.1.1 Processos de aprendizagem

No entanto, antes que um algoritmo possa aprender, ele deve ser treinado, pois o entendimento do contexto do problema e dos dados que serão utilizados são fatores que não são nada simples, mas são fundamentais para o uso correto dessas técnicas e métodos. A técnica de dividir o conjunto de dados em amostras, são úteis para avaliar a acurácia de modelos preditivos. Como deseja-se descobrir um modelo que produza previsões para dados ainda não conhecidos e com erro mínimo, a divisão faz a diferença.

O modelo é geralmente uma função matemática que tem a tarefa de encontrar padrões que melhor se aproximem ou descreva uma variável alvo com base nos dados de treinamento, por exemplo. A amostra de treino tem como objetivo responder com precisão sobre

o valor esperado da variável alvo. Na geração de previsões para os dados desconhecidos, o modelo utiliza-se do restante dos dados, ou seja, a amostra de teste que possibilita calcular o erro do modelo entre os valores observados e os valores previstos.

Com isso é possível identificar se o modelo generaliza bem os dados desconhecidos, ou seja, generalizar as funções que melhor apresentam os *outputs* desejados de forma que a mesma solução possa ser usada com outros conjuntos de dados, ao comparar o erro de previsão de vários modelos, pode-se escolher um para trabalhar de fato. E para poder generalizar a função que melhor resolve o problema os algoritmos de *machine learning* se baseiam em três passos.

A **representação**, onde o algoritmo cria um modelo que produz um resultado para um conjunto específico de *inputs*, é o conjunto de modelo que o algoritmo pode aprender, ou seja, o algoritmo cria um modelo que produz um resultado esperado a partir de dados apresentados.

Na **avaliação** o algoritmo pode criar mais de um modelo, mas não sabe a diferença de um modelo bom e um modelo ruim, a avaliação determina qual modelo funciona melhor para criar o resultado esperado. Por isso o algoritmo atribui pontuação aos modelos gerados por ele, pois mais de um modelo poderá resolver o problema.

A **otimização** refere-se ao processo de ajustar os parâmetros do modelo para melhorar seu desempenho, buscando valores de parâmetros que minimizem a diferença entre as previsões do modelo e os valores reais do conjunto de treinamento.

Existem basicamente três tipos de aprendizagem utilizados pelos algoritmos de *machine learning*, são eles:

4.1.1.1 Aprendizado supervisionado

Ocorre quando o algoritmo aprende a partir de dados de exemplos com *inputs* e possíveis *outputs* que podem conter valores quantitativos ou qualitativos à fim de prever a resposta correta quando recebe novos dados. É o termo usado sempre que um programa é treinado sobre um conjunto de dados pré-definido.

O algoritmo recebe um conjunto de entradas, juntamente com as saídas corretas correspondentes e o algoritmo aprende comparando a sua saída real com as saídas corretas. Em seguida, o algoritmo ajusta o modelo de acordo com seu processo de aprendizagem, normalmente são usados em aplicações onde dados históricos preveem eventos futuros.

O algoritmo de aprendizagem supervisionada aprende a atribuir rótulos a dados de entrada com base em exemplos rotulados fornecidos durante o treinamento. Esse algoritmo é

alimentado com o conjunto de dados de treinamento que consiste em exemplos de entrada e os rótulos correspondentes.

Ele ajusta seus parâmetros internos com base nesses dados para aprender a mapear as características de entrada para os rótulos com o objetivo de encontrar um modelo que minimize a diferença entre as previsões do modelo e os rótulos reais dos exemplos de treinamento. Após o treinamento, o modelo é avaliado usando um conjunto de dados de teste separado que não foi usado durante o treinamento, isso ajuda a verificar a capacidade do modelo de generalizar para dados não vistos.

Esse tipo de aprendizagem pode ser subdividido em dois tipos, os algoritmos de classificação são usados para o treinamento de modelos onde a saída é uma classe ou grupo à qual um dado pertence; e os algoritmos de regressão onde a saída do modelo é um valor numérico que caracteriza a relação entre as variáveis independentes e a variável alvo.

Para ilustrar isso de forma mais clara, imagine que se tem um conjunto de dados que consiste em várias amostras, cada uma com características e uma variável alvo. O objetivo do aprendizado supervisionado é aprender uma função ou mapeamento que relaciona as características das amostras à variável alvo. O modelo é treinado usando essas amostras rotuladas, e sua tarefa é prever a variável alvo para novas amostras não rotuladas que ele nunca viu antes.

4.1.1.2 Aprendizagem não supervisionada

Ocorre quando o algoritmo aprende com exemplos simples, usando dados, mas sem qualquer resposta associada, deixando a cargo do algoritmo determinar os padrões de dados por conta própria, este tipo de algoritmo tende reestruturar os dados com novos recursos que podem representar uma nova classe ou uma série de valores não correlacionados, são muito úteis a fornecer *insights* sobre o significado dos dados, ou seja, quando o algoritmo pode automaticamente encontrar relações ou padrões em um conjunto de dados.

Esse tipo de aprendizado assemelha-se aos métodos que usa-se para descobrir se certos objetos ou eventos são da mesma classe apenas observando o grau de semelhança entre eles. Os dados geralmente não possuem rótulos históricos. O algoritmo não recebe durante o treinamento os possíveis resultados, com isso, o objetivo é explorar os dados e encontrar alguma estrutura semelhante entre eles. Em vez de prever uma variável alvo, como no aprendizado supervisionado, o aprendizado não supervisionado busca identificar relações naturais entre as amostras de dados com o objetivo de organizar os dados de alguma forma ou descrever sua

estrutura.

Existem algoritmos de associação onde são capazes de identificar padrões entre os dados, como por exemplo, criar uma associação entre produtos que geralmente são comprados juntos, normalmente são utilizadas para explicar padrões entre informações independentes. Já os algoritmos de clusterização é descobrir grupos ocultos nos dados, categorizando as informações a fim de gerar segmentações e com isso agrupá-las por características parecidas.

4.1.1.3 Aprendizagem por reforço

É um aprendizado por interação, similar ao que se chama de aprender por tentativa e erro, é um paradigma avançado no campo da aprendizagem de máquina que se concentra em como agentes autônomos podem aprender a tomar decisões sequenciais para otimizar sua interação com um ambiente.

Os componentes envolvidos nesse tipo de aprendizagem são, o agente (tomador de decisão), o ambiente (onde ocorre a interação com o agente) e as ações (que são as decisões que o agente pode escolher para maximizar a premiação esperada). Tendo como objetivo fazer com que o agente escolha opções que maximizem a recompensa esperada ao longo de um determinado período de tempo.

No aprendizado por reforço o algoritmo escolhe uma ação em resposta a cada ponto de dados, recebendo uma recompensa que informa se a escolha foi boa ou não. Com base nessas informações o algoritmo muda sua estratégia para alcançar a recompensa mais alta. Nesse processo a meta é definida por meio de uma função de recompensa, que define o evento que estão ou não corretos para o agente, uma recompensa imediata define o problema enfrentado pelo agente, por isso deve ser necessariamente uma função pré-determinada.

A meta do agente é maximizar o total de recompensas recebidas, essa função a grosso modo faz o mapeamento dos estados percebidos do ambiente para um único valor, normalmente numérico, chamado de recompensa, essa função depende do ambiente e também de alguns parâmetros tais como ações tomadas pelo agente e o estado atingido pelo ambiente.

O objetivo principal da aprendizagem por reforço é encontrar a política ideal, chamada de política ótima, que maximiza a recompensa cumulativa ao longo do tempo. Durante o processo, o agente precisa equilibrar a exploração (experimentação de ações não conhecidas) e a exploração (escolha de ações conhecidas). Isso é fundamental para descobrir a política ideal. A aprendizagem por reforço é frequentemente usada em cenários em que um agente deve tomar uma série de decisões sequenciais, como em jogos, robótica, controle de sistemas autônomos,

como exemplo, os carros autônomos.

4.1.2 Aprendizagem profunda (deep learning)

Aprendizagem profunda emerge como um subdomínio da aprendizagem supervisionada envolvendo redes profundas, destacando-se pela capacidade de aprender e extrair informações complexas e hierárquicas a partir de grandes volumes de dados e reconhecer padrões que podem ser difíceis ou impossíveis de serem identificados pelo ser humano.

É uma das áreas de alto impacto da aprendizagem onde se tenta imitar o funcionamento do cérebro humano no processamento de dados e na criação de padrões para uso na tomada de decisões envolvendo uma arquitetura de redes neurais artificiais que são capazes de aprender sem supervisão humana.

Ela se concentra em treinar algoritmos conhecidos como redes neurais artificiais (RNAs) para aprender e representar automaticamente os dados em múltiplas camadas de abstração, sendo eficaz em tarefas complexas que envolvam grandes volumes de dados. Essas redes possuem camadas intermediárias, denominadas camadas ocultas, que desempenham um papel crucial na extração de características complexas dos dados.

A informação é passada por meio de cada camada, com a saída da camada anterior fornecendo entrada para a próxima camada. A primeira camada em uma rede é chamada de camada de entrada, enquanto a última é chamada de camada de saída. Todas as camadas entre as duas são referidas como camadas ocultas. Cada camada é tipicamente um algoritmo simples e uniforme contendo um tipo de função de ativação.

A representação dos dados ocorre nos diferentes níveis de abstração e a extração das características são implícitas, existindo uma elevada quantidade de parâmetros que necessitam de uma grande quantidade de dados para um aprendizado eficiente. O maior desafio no uso de aprendizagem profunda é a necessidade de grandes volumes de dados exigidos para treinar os modelos.

4.2 Observações importantes a tomada de decisão baseada em dados

Embora a tomada de decisão baseada em dados seja geralmente considerada uma abordagem confiável e informada, ela não deve ser encarada como a única fonte de informação. O aprendizado de máquina também pode ter impactos negativos que nem sempre são intencionais, levando a resultados equivocados, previsões erradas ou discriminatórias, devido a qualidade dos dados ou aplicação de um modelo impróprio para determinado problema.

Como o modelo preditivo é uma função matemática que aplicada a uma massa de dados consegue identificar padrões ocultos e prever o que poderá ocorrer, serão necessários dados de qualidade, válidos e em volume adequado para que o modelo possa aprender e consiga identificar um padrão. Como apontado, a qualidade dos dados é essencial para uma boa análise, mas a quantidade de dados também é um fator relevante, pois quando a quantidade é limitada pode-se levar a previsões tendenciosas, ou seja, se os dados de treinamento não forem representativos ou não tiverem pluralidade as previsões serão distorcidas da realidade.

Contudo, podem aparecer problemas relacionados a quantidade de dados, o primeiro deles seria o *underfitting* (sub-ajustado), quando o modelo não consegue aprender com o conjunto de dados passados, podendo levar a erros tanto no conjunto de treino como no de teste. Já o *overfitting* (sobre-ajustado) é quando o modelo aprende demais sobre o conjunto de dados de treinamento, a ponto de ajustes excessivos nos padrões encontrados nesse conjunto e não sendo capaz de generalizar bem novos conjuntos de dados.

Além disso, os dados podem fornecer informações sobre o que aconteceu no passado ou em situações similares, mas nem sempre podem prever com precisão o futuro ou capturar todos os aspectos de uma situação complexa. Sendo importante ter consciência das limitações dos dados e considerar outros fatores relevantes ao tomar decisões.

Outrossim, em determinadas situações, pode haver nuances ou elementos que os dados por si só não conseguem capturar, devendo a análise de dados ser complementada pela intuição, a experiência e o conhecimento especializado dos tomadores de decisão. Geralmente quando usados algoritmos de aprendizado de máquinas, as entradas são conhecidas e em alguns casos os resultados, mas como eles aprendem padrões a partir de dados sem implementação explícita para isso, não se sabe com exatidão quais regras são usadas por eles, podendo preservar vieses.

Eles também não podem definir valores, objetivos, prioridades e as metas específicas da organização ou da sociedade sendo necessário embasar decisões alinhadas com esses valores. E sempre deve ser considerado os aspectos éticos envolvidos, sendo fundamental garantir a privacidade, a segurança e a equidade no uso dos dados, além disso, é necessário considerar questões de justiça social para evitar possíveis vieses nos dados ou nos algoritmos utilizados.

Então, é importante destacar a responsabilidade na hora de interpretar os dados. Desenvolver pensamento crítico, com informações e conhecimento do problema a ser respondido; estabelecendo princípios éticos de conduta, uma equipe multidisciplinar e sempre

que possível, realizar questionamentos sobre os resultados. Essas são formas que podem mitigar problemas causados pelo uso de algoritmos de aprendizado de máquina e tendo em mente que nem sempre a melhor solução é a mais “tecnológica”.

5. CIDADES INTELIGENTES

Os avanços tecnológicos mudaram a vida dos homens. A facilidade e a versatilidade das comunicações permitiram quebrar barreiras como a distância e o tempo. Não importa onde se está, a internet nos traz tão perto que, em questão de segundos, parece que a pessoa com quem se comunica está do outro lado da porta.

A era digital desempenha um papel impactante na sociedade promovendo mudanças significativas na maneira como as pessoas, empresa e governos interagem entre si. Esse novo tipo de cooperação vem transformando as cidades em locais de conhecimento e informação, exigindo que o setor público tenha serviços cada vez mais orientados aos cidadãos, trazendo uma prestação mais eficiente e transparente, capazes de reduzir a distância entre a população e o setor público.

Em virtude do crescimento desordenado das cidades, surgiram vários desafios no seu planejamento, que vão desde congestionamentos de vias até o uso inadequado de recursos naturais. Neste contexto, a relevância de um planejamento urbano que contenham mecanismos mais bem informados e ágeis na tomada de decisão e que esteja alinhado aos desejos da população, são fatores determinantes da estratégia de uma administração pública moderna. E a tecnologia aliada aos dados são a chave dessa transformação.

“As cidades inteligentes caracterizam-se por uma nova forma de governar com o uso das tecnologias e conseqüente aumento da capacidade da administração pública com foco na melhoria da qualidade de vida do cidadão.” (Pereira, 2016, p.44)

Uma Cidade Inteligente e sustentável é uma cidade inovadora que utiliza as Tecnologias de Informação e Comunicação (TIC) e outros meios para melhorar a qualidade de vida, a eficiência das operações e serviços urbanos e sua competitividade, enquanto garante o atendimento das necessidades das gerações atuais e futuras com relação aos aspectos econômicos, sociais e ambientais⁶. São cidades capazes de colocar a população no centro do desenvolvimento, incorporando tecnologias na gestão urbana e utilizando esses elementos na estruturação de um governo eficiente, que engloba o planejamento colaborativo e a participação cidadã.

Seu ponto de partida deve-se a uma gestão inteligente que se beneficia do uso de tecnologias como *big data*, análises preditivas, dispositivos de IoT, conseguindo assim, fornecer um valor agregado para a gestão que as capacita na resolução dos desafios urbanos, com

⁶ International Telecommunication Union (Focus Groups on Smart Sustainable Cities, 2014)

políticas públicas mais assertivas, baseadas em dados e com capacidade analítica para responder mais prontamente a situações de emergência ou calamidade, escalonando serviços conforme a demanda e principalmente promovendo a participação ativa do cidadão na gestão pública.

Pela complexidade da administração pública, soluções tecnológicas muitas vezes são colocadas em segundo plano, seja por razões orçamentária, ausência de um quadro técnico qualificado, desconhecimento dos benefícios que a tecnologia pode oferecer ou falta de interesse das lideranças locais, esses são fatores que dificultam a viabilidade de transformação dessas cidades.

Por representar grandes transformações na gestão e na forma de interação com a população, o apoio político, principalmente da liderança local, no caso o prefeito, é um aspecto fundamental nesse processo, pois projetos de médio e longo prazo, como os de *big data*, devem ser institucionalizados e bem alinhados com o planejamento e orçamento das cidades, devendo ser visto como um projeto municipal e não da gestão atual.

As novas tecnologias trouxeram grandes mudanças no desenvolvimento da sociedade, não só no nível social e econômico, mas no jurídico, uma vez que surgiu a necessidade de regular e responder a novas questões, interesses e conflitos.

Desse modo, financiar um projeto de cidade inteligente pode ser uma tarefa complexa devido à natureza abrangente e aos custos envolvidos, entretanto, não existe a necessidade para que essas iniciativas sejam implementadas de uma só vez, podem e devem ser iniciadas com pequenas ações, onde os resultados possam ser sentidos pela população, como, por exemplo, a melhoria de algum serviço público.

5.1 O caminho para cidades inteligentes

Para que se tenha sucesso o envolvimento de todos os setores da sociedade é esperado, devendo haver uma visão compartilhada e colaborativa, com foco no bem-estar da população e deve ser executado de forma resiliente a mudanças de gestão, pois devem ser consideradas iniciativa municipal e não de uma gestão. O apoio das lideranças locais é necessário e fundamental para o sucesso desse projeto, se os líderes não estiverem convencidos da importância dessas soluções, é menos provável que forneçam o apoio necessário, fazendo com que o projeto fracasse ou nem seja cogitado.

O primeiro obstáculo é transformar o próprio modelo de funcionamento dos municípios, exigindo uma gestão integrada que supere barreiras setoriais e promova uma visão de médio e longo prazo para a cidade.

Como a abordagem de uma gestão tradicional é muito diferente em comparação a uma gestão inteligente, antes de iniciar um projeto de transformação, a gestão deve reformular algumas mudanças dos seus processos internos que reflitam em adaptações no seu planejamento e na prestação de serviços.

A mudança deve partir de cima para baixo, ou seja, os principais tomadores de decisão da administração pública devem realizar um alinhamento de expectativas de que todas as decisões devam ser respaldadas em dados.

Essa transformação implica no desenvolvimento de um plano estratégico que realize uma avaliação inicial dos problemas, a identificação de oportunidades, a seleção dos recursos tecnológicos necessários e fontes de financiamento disponíveis, analisando os benefícios esperados para os cidadãos e estabelecendo um sistema de monitoramento centrado nesses benefícios.

Esse planejamento deve ser acompanhado de indicadores que permitam monitorar e medir seu desenvolvimento e devem ser flexíveis de modo a se adaptar a mudanças no cenário tecnológico ou político.

Ressalta-se que a Internet revolucionou os conceitos clássicos de mercado, comércio, lazer, informação, consumo, mas, ao mesmo tempo, abriu novas portas e formas de arriscar a violação de direitos dignos de proteção.

A tecnologia sempre será um fator relevante na transformação das cidades inteligentes, mas não deve ser considerada como o único fator determinante, o desenvolvimento de uma mudança cultural entre os servidores públicos e colaboradores são essenciais para fortalecer o entendimento de que a transformação é uma perspectiva abrangente, devendo ser tratada com competências que vão além dos aspectos tecnológicos.

Conforme vivenciado pela população mundial, a pandemia da COVID 19 reforçou a importância da análise de dados no apoio a tomada de decisões. O combate a crise sanitária estimulou o protagonismo do uso de tecnologias pelos governos, que buscavam superar a crescente procura de serviços públicos *on-line* pela população. A digitalização desses serviços facilitou a obtenção de informações e a realização de transações sem a necessidade de deslocamentos físicos da população.

Quando bem aplicado a digitalização elimina barreiras tornando os serviços públicos mais disponíveis por parte da população residentes em áreas mais distantes ou que se encontra em circunstâncias de fragilidade. A administração pública deve sempre ter em mente que ao facilitar o acesso a serviços públicos de forma *on-line*, deve promover a inclusão digital

de toda a população, capacitando ou fornecendo meios para que todos tenham acesso fácil a esses serviços, reduzindo assim, as disparidades sociais.

Promover canais de articulação entre os vários entes da federação, juntamente com a criação de parcerias com empresas e academias, podem acelerar a implementação dessas soluções tecnológicas, pois acelera o compartilhamento e a troca de conhecimento, mas sempre tendo em mente que a tecnologia não deve ser usada somente por seu valor revolucionário, mas sim como uma melhoria na prestação de serviços à população, ou seja, não se trata apenas de implementar uma nova tecnologia por si só, mas sim como um meio de aprimorar a acessibilidade e a oferta de serviços públicos à população..

Além dos benefícios da internet, das grandes empresas de informações, têm o mesmo espaço que uma pessoa qualquer, um lugar aberto onde quem tiver acesso à internet pode se expressar por meio de textos e vídeos, sendo o lugar democrático onde milhões de pessoas têm o mesmo espaço.

5.1.1 Fatores essenciais na implantação de cidades inteligentes

Como já mencionado, o apoio político das lideranças locais é o aspecto mais importante, pois facilita a colaboração entre as várias secretárias; conflitos de competência interna ou com outros entes federativos podem ser mitigados; agiliza a alocação de recursos financeiros do orçamento municipal ou financiamento externo para o projeto; facilita a criação de parcerias com empresas e/ou academia; podem influenciar em aspectos jurídicos e regulatórios necessários em alguma etapa do projeto; e quando esse apoio é materializado e reconhecido, torna a participação cidadã mais ativa.

As cidades podem iniciar com projetos que gerem um impacto que a população possa sentir, mas com um investimento modesto. Projetos de *big data* transformam a gestão de forma transversal exigindo que a gestão forme uma equipe multidisciplinar, composta por profissionais qualificados de diversas áreas, com amplo conhecimento técnico e de gestão.

Para minimizar dificuldades que se possam ter na formação de equipes, estabelecer parcerias com o setor privado ou com a academia, podem ser uma solução, pois oferecem benefícios significativos tanto em termos de conhecimento técnico quanto ao acesso a recursos tecnológicos e financeiros. Sempre que essas parcerias tornem realidades, alguns aspectos devem ser observados e regulamentadas, como questões sobre o compartilhamento de dados, a segurança, privacidade e o uso ético desses dados.

Sob a perspectiva tecnológica, devem ser avaliados quatro aspectos, o principal é a

infraestrutura de conectividade (fibra óptica, satélite, rádio, 3G, 4G, 5G, redes sem fio) em toda a cidade e para todo cidadão, sem essa conectividade nada seria viável, pois os dados não poderiam fluir pela cidade. Um outro aspecto é a criação de interfaces de comunicação do governo com o cidadão, seja através de aplicativos móveis, plataformas web ou outros serviços on-line, sempre tendo em mente a inclusão digital.

Essa perspectiva tecnológica poderá ser ampliada com o uso de sensores podendo ser utilizados dispositivos IoT que são fundamentais em uma cidade inteligente. São eles os principais mecanismos de coleta e distribuição de dados, muitas vezes em tempo real, servindo para monitorar vários serviços públicos, como, por exemplo, gestão de trânsito, segurança, qualidade do ar, gestão de resíduos, sistemas de iluminação, ou seja, são inúmeras as variedades e utilidades desses dispositivos.

O último aspecto é um centro integrado que possa administrar, armazenar e analisar todos os dados enviados pelos dispositivos IoT ou de qualquer outra fonte, permitindo uma visão unificada de todas as áreas da cidade, acelerando assim a tomada de decisões. Mas sem uma infraestrutura de conectividade adequada todos os outros aspectos tecnológicos serão prejudicados. A aptidão de tecnologia das cidades influencia na qualidade dos serviços digitais oferecidos pelo município.

Em síntese, dê um passo pequeno, inicie com um projeto modesto, mas que seja capaz de produzir benefícios para a população. Com um escopo pequeno, torna-se mais rápido produzir os resultados e os impactos causados, facilita também a identificação de desafios tecnológicos e técnicos durante a evolução do projeto, servindo de aprendizado na elaboração de projetos maiores.

Lembrando que, a escolha do projeto deve sempre considerar um desafio urbana onde a administração pública possa intervir para levar benefícios a população, mas sempre levando em conta a infraestrutura tecnológica, a equipe técnica ou parcerias firmadas para viabilizar o projeto. A divulgação transparente dos resultados é importantíssima para garantir a confiança tanto da população quanto dos parceiros envolvidos.

Todo projeto, independente do porte, tem que ter como principal beneficiário a população, assegurando que os projetos atendam às suas necessidades. O engajamento da população só será efetiva, caso os resultados sejam os esperados. A população deve ser sempre o centro do planejamento de qualquer política pública.

5.2 O uso de *big data* nas cidades inteligentes

A utilização de *big data* pode trazer benefícios em vários setores da cidade. Com a modernização da administração pública e o uso intenso de dados, sistemas analíticos de aprendizado de máquina conseguem processar as várias fontes de coleta de dados, ampliando a compreensão dos gestores públicos sobre a dinâmica da cidade. Abrindo assim novas oportunidades para aprimorar as políticas e serviços públicos, bem como prever e responder com mais eficiência desastres ou crises, ajudando a minimizar danos a população.

Como o principal objetivo é obter clarezas para assistir os gestores na tomada de decisões, que tragam benefícios a população melhorando os serviços, pode-se exemplificar algumas áreas onde possam ser implementadas, conforme ilustrado nos parágrafos adiante.

Na área da saúde pode-se extrair estabilidade fundamentada em redes sociais, consultas médicas, prescrições e registros de sintomas de alguma doença nas unidades de saúde, por exemplo, e utilizar essas informações no monitoramento de possíveis epidemias ou surtos de doenças, possibilitando que a gestão pública adote ações para controlar de maneira precoce e preventivo as mesmas, bem como, acompanhar a ocupação de leitos otimizando a alocação de recursos e garantindo um tratamento mais eficiente aos pacientes.

Na mobilidade urbana, dados coletados em tempo real de dispositivos como sensores de tráfego, câmeras de vigilância e aplicativos de navegação servem para monitorar o fluxo de veículos, e com esses dados pode-se otimizar a gestão do tráfego da cidade, melhorando vias públicas, criando corredores exclusivos para um determinado tipo de transporte e até mesmo notificando o cidadão sobre congestionamentos e acidentes.

Um dos principais benefícios é na área de segurança, a crescente utilização de câmeras de vigilância distribuídas na cidade, oferecem informações que conseguem identificar atividades suspeitas, comportamento agressivo ou abandonos de objetos em locais públicos e até o monitoramento de aglomerações para o controle de multidões. Além de que, ao analisar dados de incidentes passados, os órgãos de segurança se qualificam na identificação de padrões criminais, áreas de maiores risco de delitos e previsões de possíveis ocorrências.

No meio ambiente, sensores podem coletar a qualidade do ar e da água trazendo benefícios tanto para saúde pública, como prevenindo a falta ou escassez de água. Padrões climáticos podem ser analisados permitindo as autoridades ações mais proativas para diminuir possíveis danos em casos de desastres naturais.

Na área da educação, fundamental em qualquer cidade, a análise de dados pode rastrear o desempenho e progresso de alunos, professores e escolas, sendo possível ser usados para avaliar a eficácia de programas educacionais, medindo o seu progresso e identificando

áreas que precisam de melhorias.

5.3 Participação cidadã

Como as ações governamentais devem ter a população no centro das atenções, nada mais natural que sejam implementadas medidas que aumentem a participação do cidadão na elaboração de políticas públicas. Mas para que isso vire realidade, os gestores públicos devem conhecer as necessidades do cidadão, criando um ambiente onde ele se sinta parte da política, sendo que sua elaboração seja compartilhada entre a gestão e a população.

Conforme Milani (2008), a introdução da participação dos usuários na produção das políticas, permite que se promova a criação de uma rede de política pública, que realize a articulação entre setores públicos, privados e organizações da sociedade civil, que tenham interesses comuns e vantajosos para a sociedade.

Não é suficiente apenas ter a vontade de envolver os cidadãos na elaboração de políticas públicas, é essencial também motivá-los a participar ativamente. Quanto mais perceberem que suas sugestões são levadas em consideração, mais inclinados a se envolverem estariam os cidadãos.

É necessário também, a promoção da inclusão digital, ou seja, as inovações que serão desenvolvidas devem ser acessíveis a todos e devem ser simples de usar. Dessa maneira será garantido que toda a população possa de fato participar de maneira efetiva no destino da cidade. A adoção de serviços *on-line* deve ser inclusiva, garantindo que todos os meios de participação popular sejam acessíveis, levando em consideração as necessidades de pessoas com deficiências, idosos e grupos marginalizados.

A administração pública deve além de motivar o interesse dos cidadãos, criar um ambiente onde possa ocorrer essa interação, seja ele físico ou virtual. A participação cidadã requer a adoção de estratégias e ferramentas para envolver efetivamente a população nas decisões municipais.

Como visto anteriormente, uma boa iniciativa é a criação de políticas de dados abertos, fundamental para a transparência e a colaboração do uso dos dados governamentais, também pode-se citar a criação de plataformas *on-line*, como websites, aplicativos móveis e redes sociais com a finalidade de facilitar o acesso a serviços disponibilizados pela prefeitura, além de conter informações sobre as políticas públicas implementadas e espaço de fóruns de discussão e debates. O orçamento participativo também é uma boa prática onde o cidadão têm a oportunidade de contribuir ativamente na decisão de como parte do orçamento público da cidade será gasta.

O orçamento participativo é um processo democrático pelo qual os cidadãos de uma comunidade têm a oportunidade de influenciar ou decidir sobre a alocação de recursos orçamentários. Nesse modelo, a população é convidada a participar ativamente na formulação, discussão e votação de propostas relacionadas ao orçamento público.

O objetivo do orçamento participativo é promover a transparência, a inclusão e a accountability no processo de tomada de decisões orçamentárias. Isso permite que as pessoas tenham um papel ativo na determinação das prioridades e necessidades locais, contribuindo para uma gestão mais eficiente e justa dos recursos públicos.

6 DESAFIOS DO BIG DATA NA GESTÃO PÚBLICA

Se a utilização de *big data* só resulta em benefícios na gestão pública, qual o motivo de não ser adotada por todos os entes da federação? Por refletir mudanças substanciais no formato dos procedimentos e processos existentes na administração, bem como na interação com o cidadão, resultando em uma mudança de paradigma, alguns desafios deveram ser enfrentados e superados pela administração pública antes de começar a caminhada em direção a uma gestão baseada em evidências.

6.1 Apoio político

Como a maioria das cidades brasileiras possuem orçamento limitado, dando prioridade em resolver problemas mais urgentes ou que tenha resultado mais imediato, permitindo assim, maior visibilidade da gestão perante a opinião pública, torna-se fundamental o apoio das lideranças locais, pois são elas que definem a visão estratégica da cidade.

A implementação de projetos de *big data* requer investimentos financeiros (apesar de atualmente ter tido reduções significativas, além da disponibilidade de várias opções *open source* para o desenvolvimento desses projetos) e de pessoal qualificado, sendo comum que os benefícios tangíveis não sejam percebidos imediatamente, necessitando de uma transição mais gradual para os resultados serem percebidos, então a alocação e priorizando de recursos orçamentários e humanos podem ser facilitados com o apoio da gestão.

A colaboração interna entre as secretárias é essencial para o desenvolvimento e implementação de políticas públicas integradas, pois facilita o compartilhamento de dados e informações na tomada de decisões. Nesse caso, o prefeito desempenha um papel fundamental na remoção de obstáculos, em caso de conflitos internos ou externos, definindo uma visão clara para o uso de dados na administração local.

A coleta, armazenamento, análise e divulgação de evidências, bem como a criação de parcerias ou contratação de pessoal, estão sujeitas habitualmente a aspectos jurídicos e regulatórios, sendo o apoio político uma maneira de facilitar a compatibilização desses projetos no aspecto legal.

Projetos associados a governos específicos correm o risco de serem descontinuados com a mudança de liderança, onde há sucessões frequentes ou disputas partidárias intensas. Uma forma de garantir a continuidade do projeto seria a sua institucionalização, tornando-o parte da estratégia da cidade, garantindo uma maior probabilidade de investimentos contínuos e facilitando a padronização de práticas realizadas ao longo das etapas do projeto. Para alcançar

a institucionalização é essencial envolver diversas partes interessadas, incluindo líderes políticos, setores da administração pública e a comunidade local.

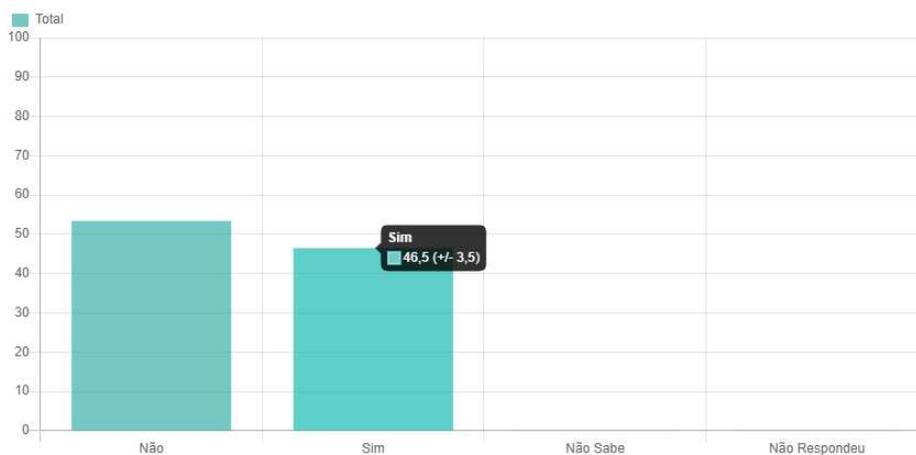
6.2 Quadro técnico limitado

Segundo Bouskela *et al.*, (2016), um elemento fundamental diz respeito às considerações relacionadas aos recursos humanos, mais precisamente à constituição de uma equipe qualificada, de natureza multidisciplinar e constantemente atualizada, apta a viabilizar um projeto de *big data*.

Como em qualquer projeto, a equipe técnica é fator fundamental para o sucesso, sem as habilidades e conhecimentos necessários nenhum projeto será viável. Projetos de *big data* exigem conhecimentos em diferentes áreas e especialidades, tornando necessário uma equipe multidisciplinar o que dificulta muito a criação da mesma, principalmente em cidades de pequeno porte.

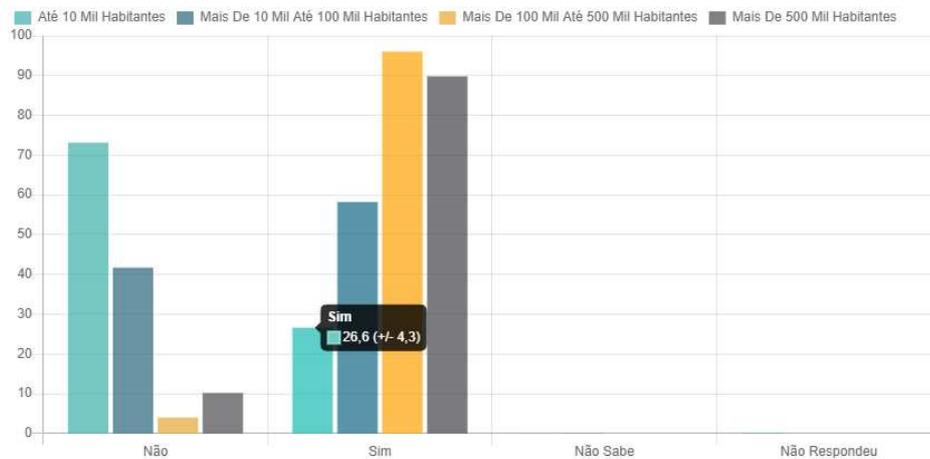
A realidade da maioria das prefeituras brasileiras reflete bem isso, menos da metade delas não possuem sequer alguma área ou departamento de tecnologia da informação, e o problema é agravado de acordo com o porte da cidade, como indica a pesquisa realizada em 2021 pela Cetic.br.

Gráfico 2 - Prefeituras com área ou departamento de tecnologia da informação - Total



Fonte: Centro Regional de Estudos para o Desenvolvimento da Sociedade da Informação (Cetic.br).

Gráfico 3 - Prefeituras com área ou departamento de tecnologia da informação – Por porte



Fonte: Centro Regional de Estudos para o Desenvolvimento da Sociedade da Informação (Cetic.br).

Uma forma de estruturar uma equipe multidisciplinar é por meio de instrumentos legais e regulatórios, onde seja desenvolvida políticas internas claras sobre o uso de dados, acesso e responsabilidades da equipe. Caso existam dificuldades na criação de uma equipe interna da prefeitura, a criação de parcerias governo-academia ou governo-empresas pode ser uma solução.

A academia possui um notável potencial no desenvolvimento de soluções inovadoras, já as empresas frequentemente têm acesso a recursos financeiros, infraestrutura tecnológica e em alguns casos projetos de *big data* já bem definidos.

6.3 Falta de integração e padronização dos dados disponíveis

Os processos baseados em dados não são nenhuma novidade no setor privado, já no setor público ainda permanece muito enraizado as práticas da gestão tradicional, onde o “achismo” era a prática comum. Em muitos casos ocorre uma resistência à mudança ou falta de compreensão sobre como os dados podem realmente otimizar processos e decisões. Como as práticas antigas da administração pública sofre com a falta de integração e padronização quando o assunto são os dados, tornando difícil uma visão precisa de onde extrair as informações necessárias.

A falta de padronização muitas vezes resulta em inconsistências dos dados, como formatos variados, nomenclatura diferente e valores ausentes, tornando a análise complexa e sujeita a erros, podendo aumentar os riscos de segurança e privacidade, tornando mais difícil

garantir a proteção adequada dos dados sensíveis. A falta de integração leva a redundância de informação que dificulta a coleta e a credibilidade em saber em qual fonte confiar.

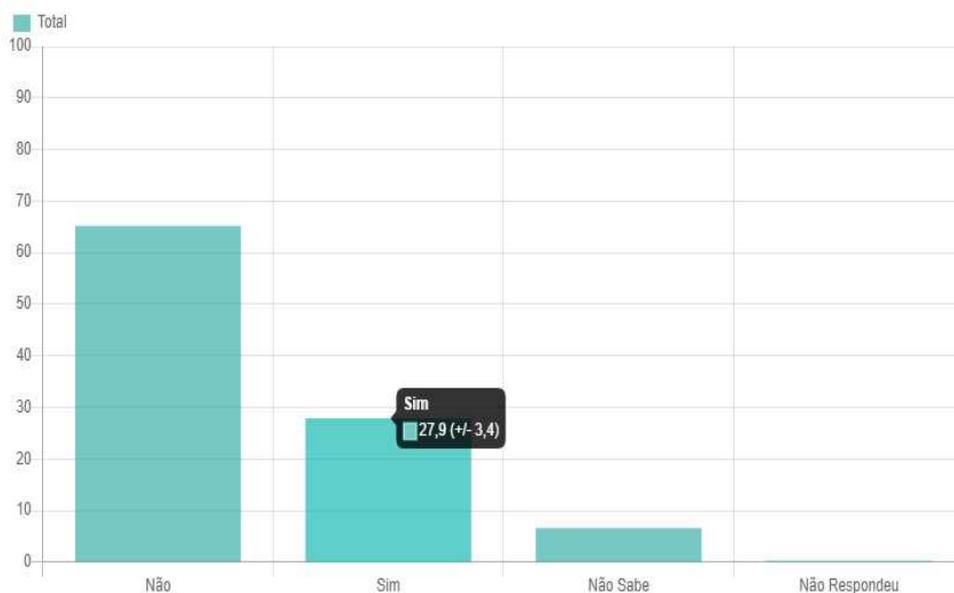
Estabelecer políticas que definam responsabilidades, procedimentos que garantam a qualidade dos dados, políticas de acesso e segurança é um grande desafio, mas também fundamental, principalmente se a política de dados abertos for implementada.

A política de dados abertos é uma boa forma para solucionar a falta de integração e padronização dos dados, pois o governo pode ter todos os seus dados publicados em um único local, facilitando a adoção de uma governança de dados e simplificando o seu uso pela gestão ou por terceiros.

A governança de dados deve estabelecer políticas bem definidas na administração dos dados em todos os níveis desde a coleta até o descarte, com definição de padrões de qualidade para esses dados, regras de privacidade, segurança da informação e conformidade com regulamentos. Conscientizado a gestão que os dados devem ser tratados como ativos estratégicos.

Apesar de ser um fator de extrema relevância a integração e padronização dos dados, essa ainda é uma prática pouco comum nas prefeituras, segunda pesquisa realizada em 2021 pela Cetic.br.

Gráfico 4 - Prefeituras, por existência de área ou pessoa responsável por procedimentos e políticas para a coleta, o armazenamento ou o uso de dados pessoais ou pela implementação da LGPD.



Fonte: Centro Regional de Estudos para o Desenvolvimento da Sociedade da Informação (Cetic.br).

6.4. Tecnológico e infraestrutura

A infraestrutura e a tecnologia desempenham um papel fundamental no sucesso de iniciativas de *big data*, muitos projetos podem fracassar quando a administração não possui uma visão clara das necessidades e do que realmente existe de concreto na cidade em termos tecnológicos.

A evolução e consolidação das tecnologias levaram a uma redução de custo, apesar disso, os governos ainda dão preferência a outras áreas. As soluções em *cloud computing* (computação em nuvem) podem ser uma estratégia eficaz para reduzir os custos de infraestrutura oferecendo uma série de benefícios que podem ser particularmente vantajosos para a administração pública que tem carência de um parque computacional para suprir as necessidades de um projeto de *big data*. Existem também muitas opções de código aberto para o desenvolvimento de soluções, como exemplos, as linguagens R e Python, o framework Apache Hadoop, bancos de dados como o PostgreSQL e o MongoDB.

A internet tornou-se uma das principais formas de acesso à informação pelas pessoas, o setor público sentiu o aumento do acesso aos seus serviços *on-line*. Com o advento da pandemia e a infraestrutura de conectividade em uma cidade, é essencial para que isso ocorra. Quando fala-se em conectividade envolve-se várias tecnologias como fibra óptica, 3G, 4G, 5G, rádio ou até mesmo satélite, apesar da ampliação dos serviços de conectividade muitas administrações municipais ainda não possuem provedores de acesso.

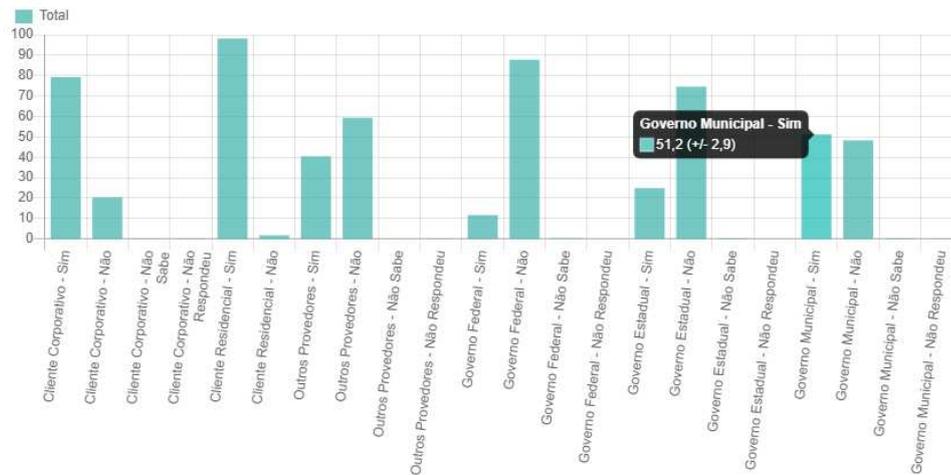
Sem conectividade nenhuma solução tecnológica se viabiliza, sendo uma prioridade de qualquer gestor municipal traçar um plano que garanta a infraestrutura necessária para que as informações circulem de maneira eficiente na cidade.

Os investimentos em infraestrutura e tecnologia de *big data* devem estar alinhados com os objetivos estratégicos do governo e precisam considerar cuidadosamente os custos associados à aquisição, implantação e manutenção desses sistemas, priorizando projetos que realmente beneficiem a administração pública e os cidadãos.

Para conduzir de maneira apropriada o setor público deve perceber a importância das tecnologias, principalmente a infraestrutura de conectividade, e priorizar o desdobramento de políticas com o apoio de um arcabouço jurídico regulatório, capazes de atender as necessidades tecnológicas que um projeto de big data possa demandar.

Infelizmente um pouco mais de 50% das prefeituras possuem contratos com provedores de internet, segundo pesquisa realizada pela Cetic.br em 2020.

Gráfico 5 - Empresas provedoras, por tipo de cliente



Fonte: Centro Regional de Estudos para o Desenvolvimento da Sociedade da Informação (Cetic.br).

6.5 Cibersegurança

A disponibilização cada vez maior de serviços *online* impulsionou um aumento na tentativa de invasões nesses serviços. Segundo relatório da *NETSCOUNT*⁷, o Brasil sofreu 328.326 ataques somente no 1º semestre de 2023, tornando a cibersegurança uma necessidade.

A administração pública deve criar mecanismos de segurança na salvaguarda de dados em sua posse, devido ao grande volume e variedade, suas bases são muito visadas, pois dispõem de dados sensíveis, como informações pessoais de servidores e cidadãos, dados financeiros da prefeitura e de terceiros, além de documentos sigilosos.

Um ataque bem-sucedido que implique no uso de informações de terceiros que comprometam a sua privacidade e segurança iria acarretar uma desconfiança nos cidadãos em relação a gestão, podendo ter repercussões políticas, sociais e legais na administração. O não cumprimento de aspectos legais de caráter obrigatório, independente do caso, também podem resultar em implicações legais e multas substanciais, além de evidenciar a falta de compromisso com a segurança perante os cidadãos aumentando a desconfiança dos serviços oferecidos pela cidade.

A identificação de responsabilidades e obrigações da equipe técnica, das políticas de governança de dados (que devem estar de acordo com as leis e regulamentos vigentes) e um inventário atualizado dos recursos computacionais, são formas de proteção contra ataques de cibersegurança em uma cidade. O conhecimento completo desses ativos é fundamental para protegê-los. Ao identificar os recursos, é possível avaliar sua vulnerabilidade em relação a

⁷ <https://www.netscout.com/threatreport/latam/brazil/>

ataques cibernéticos, além de priorizar os ativos mais críticos que precisam de medidas de segurança mais robustas.

6.6 Jurídicos e regulatórios

Além das leis de privacidade de dados como a Lei Geral de Proteção de Dados (LGPD), os projetos de *big data* podem estar sujeitos a leis e regulamentações específicas de setores, como saúde, educação e finanças. O cumprimento dessa legislação pode ser algo complexo, podendo, de acordo com a jurisdição exigir um esforço adicional para garantir que os projetos estejam em conformidade legal.

Em muitas administrações municipais existe uma carência de normas claras que auxiliem na adoção de políticas para o uso de dados dos cidadãos, devendo ser adotado pelos municípios iniciativas colaborativas e proativas que envolvam as partes interessadas, na garantia da proteção, privacidade, e segurança de dados conforme a regulamentação legal vigente.

O grande desafio dessas legislações é equilibrar a proteção da privacidade e o uso de dados pessoais por terceiros, alguns princípios devem ser estabelecidos para garantir a transparência e o uso legal desses dados. O consentimento da pessoa de forma clara para o uso de seus dados pessoais, esclarecendo a finalidade específica onde o dado será usado é uma forma de promover o uso responsável dos dados pessoais por terceiros.

6.7 Considerações éticas relacionados à privacidade e proteção de dados pessoais

No contexto de *big data*, a análise proporciona uma relação mais penetrante nos dados e não ocorrendo de forma transparente sobre como são coletados, cria-se uma vulnerabilidade do cidadão em relação a administração que detém a tutela dos dados, implicando em preocupações éticas da utilização dos mesmos.

Quando não existe o consentimento expresso do titular do dado pessoal sobre a utilização e necessidade do seu uso, cria-se um risco ao cidadão e em casos mais extremos a própria democracia.

A ética na coleta de dados é fundamental na garantia de que os dados serão usados de maneira justa e responsável, sem a preocupação de se produzir um ambiente que resulte do abuso no uso de informações privilegiadas, sendo necessário garantir que os cidadãos estejam cientes de como seus dados serão coletados, processados e utilizados para fins de análise, permitindo que os indivíduos tenham controle sobre suas informações e que sejam oferecidas

opções claras de consentimento.

Em algoritmos de aprendizado de máquinas, as análises preditivas podem encontrar correlações nos dados, mesmo nos casos em que não exista causa e efeito diretos entre dois fenômenos. Nos casos onde essas análises são aplicadas a uma parcela da população, conclusões injustas ou discriminatórias podem surgir, causando consequências negativas sobre algum cidadão ou parte da população.

O viés discriminatório dos resultados das análises pode ter impactos prejudiciais para a sociedade e levam gestores a tomar decisões baseadas em conclusões irreais. A confiança cega nos resultados pode resultar em decisões com implicações éticas e de privacidade, a análise de dados é uma ferramenta poderosa, porém cercada de riscos. Para Buttareli (2015), a excessiva confiança nas previsões estatísticas pode fazer com que a análise de dados se torne uma ferramenta que promove a discriminação.

A carência de dados de alta qualidade pode gerar previsões discriminatórias ou injustas, dados históricos coletados com vieses discriminações, caso sejam usados, refletiram esses vieses. Problemas também podem aparecer durante o processo de seleção e preparação dos dados, onde decisões sobre quais dados incluir ou excluir podem influenciar em resultados tendenciosos.

Para mitigar o viés nos projetos de *big data*, é essencial adotar medidas proativas que realizem avaliações nos dados antes da análise para identificar e compreender qualquer tendência ou desigualdade, montando equipes de análise de dados que sejam diversas em termos de experiência, origem e perspectivas, podem reduzir o viés na seleção e interpretação dos dados. A administração pública deve ter o compromisso moral e ético na hora de explicar os dados.

Ter transparência sobre os dados usados, as metodologias de análise e os resultados obtidos é essencial para construir um relacionamento de confiança que promova condutas éticas e rotinas de proteção à privacidade dos dados. Ela serve como um mecanismo de proteção dos direitos dos cidadãos, garantindo que seus dados sejam usados de maneira ética e de acordo com a lei, desempenhando um mecanismo de confiança dos cidadãos com a administração pública.

7 ETAPAS PARA A EXECUÇÃO DE UM PROJETO DE BIG DATA

Os projetos de *big data* têm o potencial de transformar significativamente a forma como os governos operam e prestam serviços à população. Iniciar um projeto piloto de *big data*, requer uma abordagem cuidadosa e estratégica para garantir o sucesso e o valor agregado. Projetos como esse podem ser complexos e desafiadores, mas quando bem planejados e executados, podem gerar evidências valiosas e melhorias significativas nos serviços públicos.

7.1 Identificação de objetivos e desafios

A identificação e desafios que a gestão pública enfrenta, priorizando-os com base na sua importância e impacto na comunidade, desenvolve critérios claros para priorizar os desafios e é fundamental, podendo incluir a gravidade do problema, o impacto na qualidade de vida dos cidadãos, a viabilidade de soluções, o custo de implementação e a urgência.

Caso sejam observados mais de um desafio, é importante priorizá-los com base em sua importância e impacto potencial. Determinando quais objetivos são de alta prioridade e quais podem ser abordados em fases posteriores do projeto, se atribue pontuações a cada desafio com base nos critérios estabelecidos pela gestão, e isso permite classificar os problemas em ordem de prioridade.

O uso de técnicas como a matriz de priorização ou a técnica dos "5 Porquês", idealizada pelo arquiteto do sistema Toyota de produção, Taiichi Ohno, é uma abordagem de resolução de problemas que busca identificar a causa raiz de um problema ou desafio, fazendo repetidamente a pergunta "Por quê?" até chegar à causa fundamental. Essa técnica pode ser adaptada para explorar objetivos e desafios de forma mais profunda, convertendo os desafios identificados em objetivos específicos e mensuráveis.

Um diagnóstico geral, reflete sobre o(s) problema(s) encontrado(s) do ponto de vista do desafio, qual a solução ou possíveis soluções, como poderá ser medido o desempenho do projeto e qual seria o desempenho mínimo necessário para atingir o objetivo, pense na tecnologia e infraestrutura envolvida e a equipe técnica requerida para o desenvolvimento da solução.

Os objetivos definidos, as prioridades e os indicadores de sucesso em um plano formal, servem como um guia durante todo o projeto e ajuda a manter o foco nas metas, envolvendo as partes interessadas, ajuda a garantir que os objetivos estejam alinhados com as necessidades reais e que haja apoio desde o início do projeto.

7.2 Formar uma equipe qualificada

É importante identificar as habilidades e competências necessárias para o projeto para montar uma equipe multidisciplinar composta por membros com formação em diferentes áreas de especialização, definindo papéis e responsabilidades claros para cada membro da equipe.

A criação de parcerias com empresas e instituições acadêmicas podem ser fatores fundamentais. Além da capacidade técnica e tecnológica que elas possuem, poderá ser uma alternativa em casos de dificuldades em montar uma equipe técnica. Sendo que em qualquer dos casos, é necessário identificar as responsabilidades de todas as partes envolvidas e os interesses comuns devem ser alinhados com os objetivos do projeto.

Se esclarece que as equipes de trabalho entre pessoas localizadas em lugares diferentes, que criam um projeto de trabalho comum, caracterizado por uma autonomia de trabalho, colaboração de dados e conhecimento, interdependência e por um conceito de seu trabalho, fundamentado nos resultados produzidos.

Além disso, as questões relacionadas à propriedade intelectual, à segurança de dados e à conformidade regulatória também devem ser tratadas adequadamente em acordos de parceria. A designação de um líder experiente, para supervisionar e coordenar as atividades da equipe, poderá ser importante na condução do projeto, facilitando a tomada de decisões e comunicação com as partes interessadas, além de fornecer recursos, apoio político e removendo obstáculos quando necessário.

7.3 Avaliar os recursos existentes

Assim como ter uma equipe técnica, o levantamento e avaliação da capacidade dos recursos computacionais e de infraestrutura de conectividade existentes para estabelecer se são adequados para sustentar as necessidades do projeto é de extrema importância.

Com base no levantamento, possíveis gargalos ou limitações podem ser identificados. Deve-se fazer um plano para atualizar ou expandir os recursos de TIC conforme necessário, estimando os custos associados às atualizações ou expansões necessárias.

A contratação de provedores de nuvem pode ser uma decisão estratégica importante nessa etapa, pois oferecem serviços de infraestrutura, armazenamento de dados e recursos de computação que podem ser essenciais para a implementação e operação bem-sucedida do projeto, levando em consideração os requisitos específicos, os custos associados e as considerações de segurança e conformidade.

Apesar de óbvio, sem uma infraestrutura de conectividade eficiente nas cidades a migração para computação em nuvem poderá ser inviável, pois os recursos e serviços dos provedores em nuvem enfrentarão dificuldades para serem acessados.

No caso de uma migração para computação em nuvem os gastos ficariam variados de acordo com os recursos utilizados em relação a processamento, armazenamento, tempo de uso, dados utilizados e em alguns casos o suporte técnico. Também pode ocorrer uma mudança na estrutura da equipe de TI, devido ao provedor do serviço em nuvem assumir a responsabilidade por alguns recursos, principalmente os relacionados a administração da infraestrutura.

Em governos locais, principalmente os de pequeno porte, o número reduzido de usuários que acessam os serviços on-line da prefeitura, em muitos casos, não é suficiente para justificar a aquisição e manutenção de uma infraestrutura de TIC própria, podendo a computação em nuvem ser uma alternativa viável e econômica ao município.

Uma equipe técnica qualificada que consiga implementar e gerenciar os recursos oferecidos pelos provedores é algo necessário, mas essa equipe também deve ter o conhecimento de conseguir mapear o que o modelo de negócio precisa e qual recurso ou serviço em nuvem suprirá a necessidade. Só assim a administração pública poderia se beneficiar dos recursos e serviços com a possibilidade de redução de despesas potenciais.

Um grande problema encontrado para a utilização em computação em nuvem é a ausência de regulamentação. Esses entraves normativos criam insegurança em relação a viabilidade de contratação de provedores devido as incertezas nas questões relacionadas a modalidade de adoção, precauções necessárias, riscos associados, entre outros.

7.4 Diagnóstico dos dados

Após os levantamentos iniciais, é hora de listar os dados que serão necessários para o projeto, isso pode incluir dados governamentais, dados de cidadãos, dados de sensores, entre outros. Obtenha acesso em dados pertencentes a terceiros e verifique as recomendações legais necessárias para obter autorização de acesso.

Analise a qualidade dos dados em cada fonte, considerando fatores como precisão, integridade, consistência, atualidade, confiabilidade dos dados e também problemas comuns, como dados ausentes, duplicatas e erros. Identificando assim a existência de lacunas significativas nos dados que possam afetar a qualidade da análise, incluindo a identificação de variáveis ausentes ou informações não registradas. A qualidade dos dados é fundamental, e é

necessário garantir que os dados sejam coletados de forma ética e estejam em conformidade com as regulamentações de privacidade.

Crie um mapeamento dos dados que descreva as relações entre os diferentes conjuntos de dados, isso facilitará o rastreamento em caso de análises futuras, pois ajuda a entender como os dados estão interconectados e a identificar pontos de integração ou junção, mantendo sempre uma documentação detalhada sobre os dados disponíveis, incluindo a origem, o formato, a qualidade e qualquer transformação realizada.

7.5 Garantir a privacidade e a segurança dos dados

Certifique-se de cumprir regulamentações de privacidade, como a Lei Geral de Proteção de Dados (LGPD), a privacidade e a segurança dos dados são preocupações críticas no setor público, identifique os dados sensíveis que serão usados no projeto, isso inclui informações pessoais, informações confidenciais do governo e qualquer outro tipo de dados que precise de proteção especial, e planeje medidas de segurança de dados apropriadas para protegê-los.

Caso seja necessário exclua ou proteja os dados confidenciais, considerando sempre que possível a anonimização ou pseudonimização de dados pessoais (envolve a remoção de informações identificáveis ou a substituição delas por identificadores exclusivos para proteger a privacidade dos indivíduos).

Realize controles de acesso que garantam que apenas pessoas autorizadas tenham acesso aos dados sensíveis, e efetue uma avaliação de possíveis riscos de segurança no controle de acesso aos dados, identificando possíveis ameaças e vulnerabilidades.

7.6 Preparação dos dados

Desenvolva um plano de limpeza e preparação de dados, envolvendo a correção de erros, corrija ou remova *outliers*, preencha valores ausentes ou elimine suas linhas, elimine atributos que não forneçam informações úteis, preencha lacunas, padronização formatos e outras ações para melhorar a qualidade dos dados.

A preparação de dados é uma atividade iterativa e é a etapa que consome mais tempo em projetos de *big data*. Essa etapa garantirá a qualidade, confiabilidade, padronização e consistência dos dados, onde valores ausentes foram identificados e tratados, *outliers* foram identificados e retirados, garantir que os dados estejam prontos para serem alimentados em modelos de forma eficiente.

A qualidade dos dados preparados desempenha um papel crítico no sucesso de qualquer análise ou modelo de aprendizado de máquina, por isso é importante investir tempo e esforço nessa etapa.

7.7 Análise de dados

A análise de dados é o ponto chave de um projeto de *big data*, é onde ocorre a mudança dos dados em informações úteis, trazendo conhecimentos valiosos que podem ser usados pela administração pública em tomadas de decisões, identificação de tendências e anseios da população, além de antecipar ações previstas em padrões contidos nos dados (sendo ocultos ou não).

“Na era do big data, a tomada de decisões sobre quais variáveis examinar com base apenas em hipóteses não é mais eficiente. Os bancos de dados são grandes demais, e a área sob investigação, provavelmente complexa demais. Por sorte, muitas das limitações que nos obrigaram a usar a abordagem com base em hipótese já não existem da mesma forma.” Schönberger-Mayer e Cukier (2013, p.04)

Explore os diferentes modelos e algoritmos de aprendizado de máquinas disponíveis para resolução do problema. Utilizando o máximo de dados possíveis, lembrando em dividi-los em conjuntos que possam ser usados no treinamento e no teste dos modelos. Analise as variáveis mais significativas e os tipos de erro que cada modelo possa cometer, avaliando e comparando o desempenho de cada modelo.

7.8 Desenvolva a aplicação

A escolha de estratégias e ferramentas deve ser adaptada aos requisitos específicos do projeto, lembrando também que todo o desenvolvimento pode ser realizado em nuvem. Manter-se atualizado com as melhores práticas e tendências em desenvolvimento de software para garantir que sua aplicação seja eficiente, segura e escalável.

Atualmente existem diversas linguagens de programação open source disponíveis para o desenvolvimento de projetos de big data (R, Python, Java, PHP, entre outras). Uma iniciativa do Banco Interamericano de Desenvolvimento ⁸(BID), por exemplo, disponibiliza várias ferramentas de código aberto, que são bem documentadas e altamente reutilizáveis, com uma avaliação da qualidade do código e uma licença que permite a reutilização e adaptação da ferramenta.

⁸ <https://code.iadb.org/pt>

7.9 Apresentação e comunicação

Após a implementação do projeto, deverá ser apresentado o resultado de maneira eficaz aos tomadores de decisão e ao público em geral. Uma documentação contendo o que foi feito, explicando como os objetivos do desafio foram atingidos, esmiuçando a natureza do desafio confrontado, dimensionando sua magnitude e delineando seu escopo, identificando os indivíduos impactados e o custo do problema e as opções de cada ação e seus impactos. A documentação deve ainda enumerar as descobertas obtidas ao longo do processo, identificar elementos exitosos e malsucedidos, bem como elucidar as limitações inerentes ao projeto.

O ambiente de negócios global, a volatilidade e velocidade com que as inovações, novas tecnologias da informação levou a comunicação de forma transparente a todas as partes interessadas, inclusive a comunidade, explicando o raciocínio por trás das escolhas, ajuda na construção de confiança e apoio para as iniciativas futuras.

7.10 Acompanhamento e aperfeiçoamento contínuo

O acompanhamento de desempenho permite avaliar o impacto das políticas públicas e programas governamentais, facilitando ajustes e melhorias com base em resultados reais, além de promover a transparência, permitindo que os cidadãos tenham acesso a informações sobre como os recursos públicos estão sendo utilizados, aumentando a responsabilidade das autoridades públicas.

Realizar o acompanhamento de desempenho ajuda a ajustar a estratégia à medida que novas evidências são descobertas, executar o processo de *big data* sem questionamentos relevantes que efetivamente contribuam para o benefício do negócio não é uma abordagem viável. Projetos de big data modernizam a administração pública no processo de tomada de decisões.

7.11 Provedor em nuvem

Apesar de não ser uma etapa necessária, os provedores de nuvem podem ser uma solução viável para muitas administrações locais, pois oferecem infraestrutura escalável que permite aumentar ou diminuir os recursos de acordo com as necessidades de cada cidade, sendo particularmente valioso em projetos de *big data*, onde os requisitos de processamento e armazenamento podem variar significativamente. Lembre-se que em cidades de pequeno porte, a quantidade de usuários dos serviços on-line da prefeitura quase sempre não justifica a

aquisição e manutenção de uma infraestrutura de TIC própria.

Serviços de computação em nuvem (*cloud computing*, em inglês), referem-se a uma variedade de recursos de computação, armazenamento, rede e software disponibilizados pela internet por meio de provedores de nuvem. Em vez de adquirir e manter infraestrutura de TIC localmente, os órgãos públicos podem usar esses serviços em nuvem para atender às suas necessidades em projetos de *big data*.

A computação em nuvem oferece várias vantagens, incluindo escalabilidade, flexibilidade, economia de custos e acessibilidade. Onde os recursos contratados são automaticamente ajustáveis a demanda do cliente. A computação em nuvem pode ser dividida em diferentes modelos de implantação, que definem como os recursos são compartilhados e acessados pelos usuários. São elas:

- a) nuvem pública: infraestrutura de nuvem onde os recursos de computação, armazenamento e rede são compartilhados entre várias organizações ou indivíduos.
- b) nuvem privada: infraestrutura de nuvem dedicada a uma única organização ou empresa, oferecendo maior controle, segurança e personalização dos recursos, permitindo que a organização atenda a requisitos específicos de conformidade e segurança;
- c) nuvem híbrida: combinação de nuvem pública e nuvem privada que permite que os dados e aplicativos sejam compartilhados entre elas, as organizações podem mover cargas de trabalho entre nuvens públicas e privadas de acordo com as necessidades, proporcionando flexibilidade e escalabilidade;
- d) nuvem comunitária: infraestrutura de nuvem compartilhada por várias organizações com interesses comuns ou requisitos de conformidade, permitindo que organizações com necessidades semelhantes colaborem e compartilhem recursos em uma nuvem dedicada a uma comunidade específica, oferece controle e segurança compartilhados, ao mesmo tempo em que atende aos requisitos exclusivos da comunidade, é uma alternativa interessante que pode ser usada por organizações governamentais.

Cada modelo de nuvem tem suas vantagens e desvantagens, e a escolha depende das necessidades e objetivos da administração pública, que pode optar por uma única abordagem ou contratar uma solução com abordagem híbrida ou comunitária, pois são capazes de combinar a simplicidade de uma nuvem pública com a customização e controle de uma

nuvem privada.

A Gartner, reportou em seu relatório anual de 2022, as empresas líderes do mercado de infraestrutura em nuvem e serviços de plataforma, em quatro quadrantes (líderes, visionárias, fornecedoras de nicho e desafiantes), chamado de quadrante mágico, essas empresas representam em sua categoria soluções altamente automatizadas e padronizadas, naquele ano a vendedora foi a *Amazon Web Services (AWS)*.

Figura 1 – Quadrante mágico Gartner



Fonte: Gartner

A AWS é um dos principais provedores de serviços de computação em nuvem do mundo, oferecendo uma ampla gama de serviços e recursos para atender às necessidades individuais, oferecendo desde hospedagem simples de sites até a execução de aplicativos complexos e o processamento de grandes volumes de dados.

Tabela 2 – Alguns serviços da AWS disponibilizados

Recursos	Serviços AWS
Computação em Nuvem	Amazon EC2 (<i>Elastic Compute Cloud</i>): Permite a criação de máquinas virtuais escaláveis para executar aplicativos; AWS Lambda: Permite a execução de código sem a necessidade de provisionar servidores.
Armazenamento em Nuvem	Amazon S3 (<i>Simple Storage Service</i>):

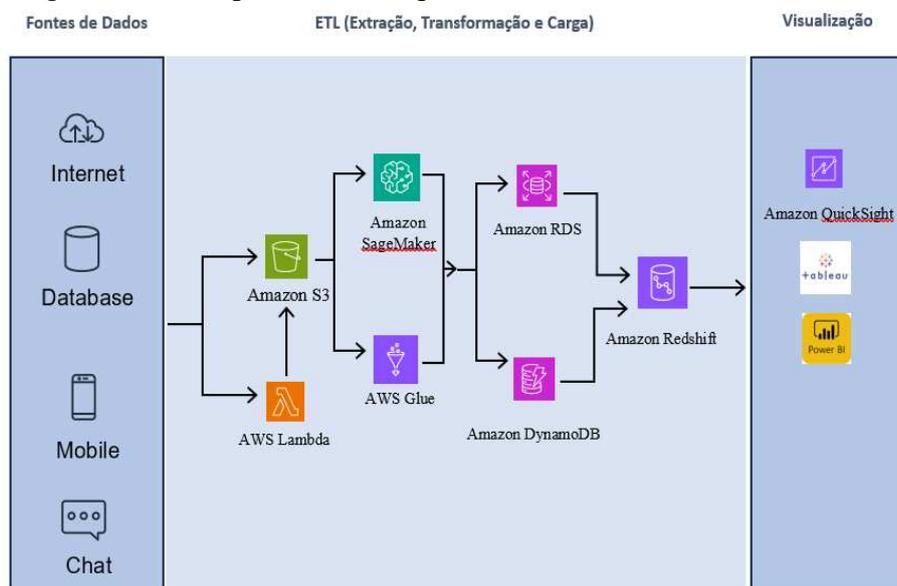
	<p>Oferece armazenamento de objetos altamente escalável;</p> <p>Amazon EBS (<i>Elastic Block Store</i>): Fornece armazenamento de blocos para uso com instâncias EC2.</p>
Banco de Dados	<p>Amazon RDS (<i>Relational Database Service</i>): Oferece bancos de dados relacionais gerenciados;</p> <p>Amazon <i>DynamoDB</i>: É um banco de dados <i>NoSQL</i> totalmente gerenciado.</p>
Rede e Conteúdo Distribuído (<i>Networking and Content Delivery</i>)	<p>Amazon VPC (<i>Virtual Private Cloud</i>): Permite a criação de redes virtuais isoladas na AWS;</p> <p>Amazon <i>CloudFront</i>: É uma CDN (<i>Content Delivery Network</i>) que ajuda a entregar conteúdo de maneira rápida e segura.</p>
Análise de Dados (<i>Analytics</i>)	<p>Amazon EMR (<i>Elastic MapReduce</i>): Facilita o processamento de grandes volumes de dados;</p> <p>Amazon <i>Redshift</i>: É um serviço de <i>data warehousing</i>.</p>
Inteligência Artificial e <i>Machine Learning</i> (IA/ML)	<p>Amazon <i>SageMaker</i>: Facilita a construção, treinamento e implantação de modelos de aprendizado de máquina;</p> <p>AWS <i>Rekognition</i>: Oferece recursos de visão computacional.</p>
Segurança e Identidade	<p>AWS IAM (<i>Identity and Access Management</i>): Permite o gerenciamento de identidades e acesso aos recursos da AWS;</p> <p>AWS Cognito: Facilita a autenticação e autorização de usuários.</p>
Internet das Coisas	<p>AWS <i>IoT Core</i>: Permite a conexão de</p>

	dispositivos <i>IoT</i> à nuvem e o gerenciamento desses dispositivos.
Serviços de ETL (Extração, transformação e carregamento de dados)	AWS Glue: Permite facilitar a descoberta, preparação, movimentação e integração de dados de várias fontes para análise.
Serviços de DevOps	AWS <i>CodePipeline</i> e AWS <i>CodeBuild</i> : São serviços de entrega contínua e compilação de código; AWS <i>Elastic Beanstalk</i> : Facilita a implantação e gerenciamento de aplicativos.

Fonte: Elaborado pelo autor.

Esses são apenas alguns exemplos dos serviços oferecidos pela AWS⁹, lembrando que são oferecidos mais de 200 serviços, muitos deles gratuitos ou liberados no período de 12 meses para uso sem custos, que pode ser usado para estudos ou criação de um projeto-piloto pequeno em *big data*. A plataforma continua a evoluir e adicionar novos serviços, para atender às necessidades de diferentes tipos de cargas de trabalho e organizações.

Figura 2 – Exemplo de uma arquitetura AWS.



Fonte: Elaborada pelo autor.

É importante entender e planejar o uso dos serviços da AWS, principalmente em relação aos modelos de cobrança, pois isso ajudará a otimizar os custos de acordo com as

⁹ Fonte: <https://aws.amazon.com/pt>

necessidades específicas, ela oferece vários tipos de cobrança de acordo com o serviço utilizado.

Cada serviço da AWS tem seu próprio modelo de preços, e a forma como é realizada a cobrança pode depender de vários fatores, como a região (localidade) em que o serviço é usado, o tipo de instância ou recurso utilizado e a quantidade de uso, ainda podem variar de acordo com o tempo de execução medido em hora, minuto ou segundos, outros são cobrados com base no número de requisições ou eventos processados, além de cobranças por armazenamento utilizado ou pela quantidade de solicitações de acesso aos dados.

A AWS também oferece um conjunto de ferramentas para monitorar e gerenciar os custos que auxiliam os clientes a controlar e otimizar seus gastos na plataforma, como por exemplo, a *AWS Pricing Calculator*, que é uma calculadora de preços *online* que permite que os clientes estimem os custos dos serviços da AWS com base em suas necessidades específicas.

A outra é o *AWS Budgets*, onde os clientes podem criar orçamentos personalizados com base em suas necessidades e metas de gastos, especificando o valor máximo que desejam gastar em um período específico, como um mês. Assim, a ferramenta permite o acompanhamento dos gastos em relação ao orçamento definido, exibindo gráficos e informações sobre como os gastos estão evoluindo ao longo do tempo, além de enviar alertas para notificar os clientes quando os gastos atingirem um determinado percentual do orçamento.

A escolha de um provedor de nuvem deve ser feita cuidadosamente, existem várias no mercado, mas deve ser sempre levado em consideração os requisitos específicos de cada projeto, os custos e as considerações de segurança e conformidade. Lembrando que uma solução em nuvem não irá ser viável se a cidade não dispor de uma infraestrutura de conectividade que satisfaça as demandas de acesso com o provedor em nuvem, além de possuir uma equipe técnica qualificada que consiga implementar e gerenciar os recursos oferecidos pelos provedores.

8. ESTUDOS DE CASOS

As Organizações das Nações Unidas, durante a Cúpula das Nações Unidas sobre o Desenvolvimento Sustentável, estabeleceu a Agenda 2030, que é um plano de ação global adotado pelos Estados-membros das Nações Unidas composta por objetivos de desenvolvimento sustentável (ODS), destinados a abordar uma ampla gama de desafios globais, incluindo pobreza, desigualdade, mudança climática, degradação ambiental, paz e justiça. Embora não haja um ODS específico dedicado aos dados, a importância dos dados é mencionada em vários como uma ferramenta crítica para a tomada de decisões informadas.

Será apresentado três estudos de caso, desenvolvidos pelo Instituto de Planejamento de Fortaleza (IPALNFOR), uma autarquia da prefeitura de Fortaleza. O primeiro deles são referentes a dados abertos georreferenciados (Fortaleza Em Mapas), já o segundo é um projeto de mapas colaborativos com a participação da população (Mapa Colaborativo), e por último, o BigData Fortaleza que é uma ferramenta de políticas públicas baseadas em dados.

A escolha desses três casos representa um diagnóstico de como um órgão relativamente pequeno, da Prefeitura de Fortaleza, conseguiu pôr em produção projetos baseados em dados. Vale ressaltar que, “Fortaleza Em Mapas” e “Mapa Colaborativo”, foram projetos desenvolvidos pela equipe interna do órgão e no seu desenvolvimento, apenas ferramentas *open source* foram utilizadas.

Já o projeto BigData Fortaleza representa uma solução mais robusta, que envolve técnicas de aprendizado de máquina e computação em nuvem. Esse projeto contou com o apoio o apoio político da prefeitura, onde firmaram parcerias com a Fundação Cearense de Apoio ao Desenvolvimento Científico e Tecnológico e a Universidade Federal do Ceará, extremamente importante, além disso, foi utilizado a AWS como provedor em nuvem.

8.1 Fortaleza Em Mapas¹⁰

A plataforma Fortaleza em Mapas foi criada com o objetivo de disponibilizar informações georreferenciadas produzidas principalmente pelo poder público e assim proporcionar ao cidadão o acesso à informação sobre a cidade de Fortaleza. Atualmente conta com mais de 400 geoserviços criados e mapas disponibilizados para a população.

Figura 3 - Quantitativo de mapas por categoria.

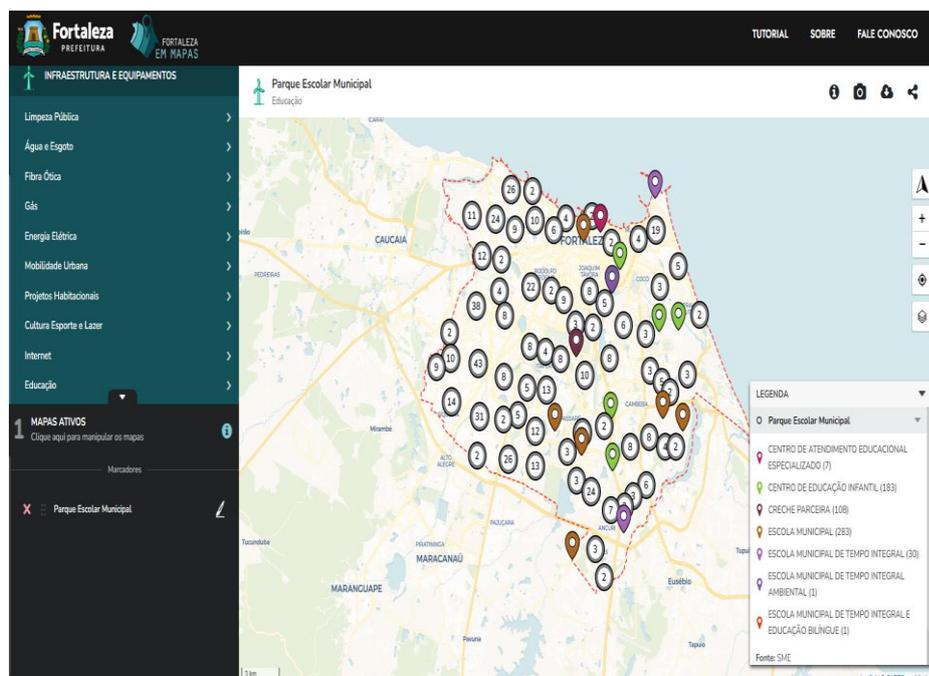
¹⁰ <https://mapas.fortaleza.ce.gov.br/>



Fonte: IPLANFOR

Seus objetivos são o compartilhamento de mapas na administração pública, atualmente contém dados de várias órgãos e secretarias, como o Instituto Brasileiro de Geografia e Estatística (IBGE), Companhia de Água e Esgoto do Ceará (CAGECE), Secretária de Finanças (SEFIN), Secretária de Saúde (SMS), Secretária de Educação (SME), dentre outras; garantir interoperabilidade de plataformas; garantir e promover o acesso à informação; e auxiliar a tomada de decisão.

Figura 4 - Localização das escolas municipais de Fortaleza.



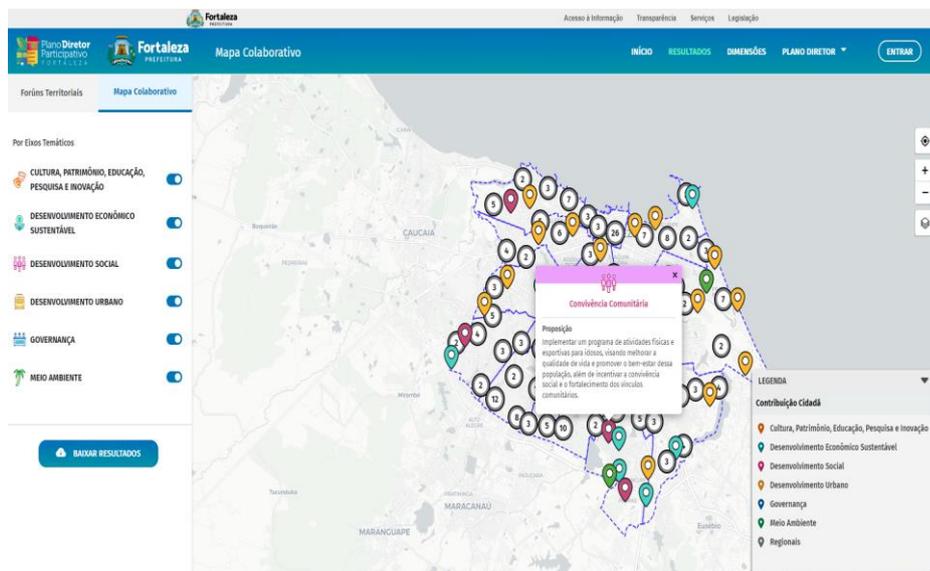
Fonte: <https://mapas.fortaleza.ce.gov.br/>

8.2 Mapa Colaborativo¹¹

É uma plataforma onde a população registra propostas que sejam de seu interesse de forma georreferenciada, incentivando a colaboração na construção do Plano Diretor Participativo (PDP), para que o planejamento do município atenda às necessidades coletivas de toda a população.

Existe a necessidade de um cadastro simplificado para propor as propostas, sendo possível visualizar propostas já realizadas, navegar por palavras chaves. As propostas são divididas em seis temas, cultura, patrimônio, educação, pesquisa e inovação; desenvolvimento econômico sustentável; desenvolvimento social; desenvolvimento urbano; governança; meio ambiente.

Figura 5 - Mapa colaborativo.



Fonte: <https://mapapdpfor.fortaleza.ce.gov.br/>

8.3 BigData Fortaleza¹²

Ficou evidente a relevância da implementação de políticas públicas baseadas em dados para auxiliar a administração pública na formulação de decisões informadas que promovem a evolução de estratégias eficazes para a realização de seus propósitos.

Com base nisso, o IPALNFOR, firmou parcerias por meio do programa Cientista Chefe, que busca trazer inovação para a gestão pública, através do trabalho cooperativo entre cientistas e gestores públicos, com a Fundação Cearense de Apoio ao Desenvolvimento

¹¹ <https://mapapdpfor.fortaleza.ce.gov.br/>

¹² <https://bigdata.fortaleza.ce.gov.br/>

Científico e Tecnológico (FUNCAP) e o Grupo de Pesquisa, Desenvolvimento e Inovação em Redes de Computadores, Engenharia de Software e Sistemas (GREat) vinculado à Universidade Federal do Ceará (UFC), que atua em conjunto com diferentes cursos e departamentos da universidade, compondo uma equipe multidisciplinar, para o desenvolvimento da plataforma BigData Fortaleza.

A plataforma integra dados de diferentes secretarias municipais para a visualização customizada de informações, análise de cenários e tomada de decisões baseadas em evidências, e integrando diferentes fontes de dados que permite a formulação de novos analíticos.

Gestores podem definir e receber alertas específicos criados a partir de indicadores disponíveis na plataforma, personalização a visualização de dados e de notificações, e assim, sendo possível direcionar ações públicas de acordo com as informações recebidas.

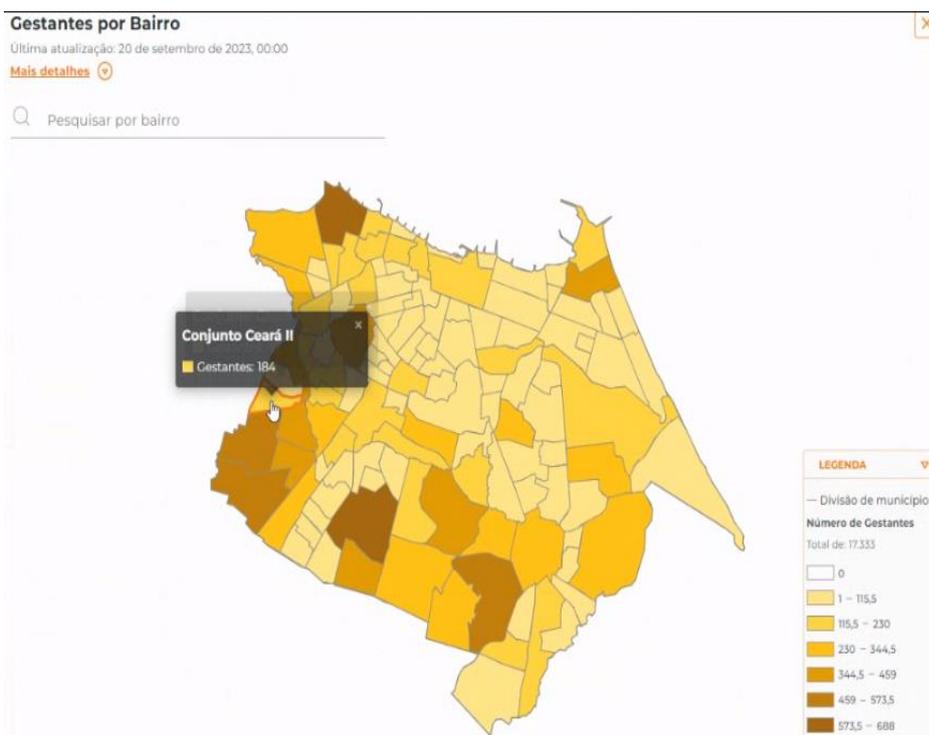
O projeto piloto dessa ferramenta teve como domínio a primeira infância, que contou com a parceria do IPLANFOR com a Coordenadoria Especial da Primeira Infância (CESPI), a Secretaria Municipal da Educação (SME) e a Secretaria Municipal de Saúde (SMS).

Visa investigar o desafio da demanda de creches na cidade de Fortaleza a fim de subsidiar os gestores no processo de tomada de decisão a partir de dados, realizando o levantamento necessário para auxiliar na compreensão do contexto das crianças que precisam de creche, estimando indicadores do Plano Municipal de Educação (P.M.E.) relacionados à etapa creche e propiciando dados estratégicos para ampliação da oferta de vagas.

Buscou-se entender a distribuição da população de 0-3 anos por bairro de forma atualizada, com essas informações os gestores municipais conseguem, por exemplo, identificar quais creches precisarão de novas vagas.

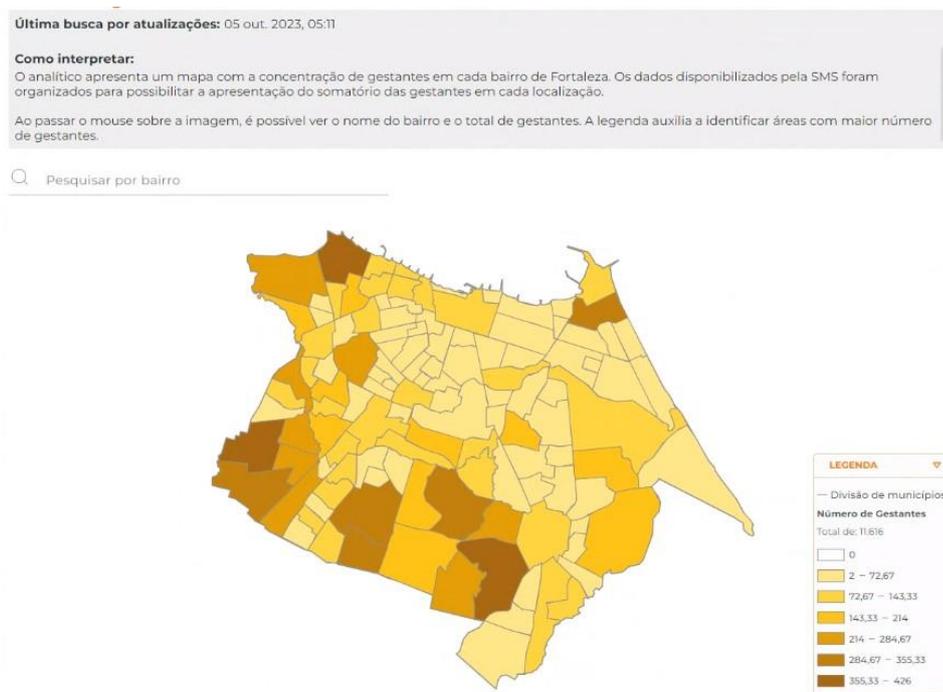
Como a plataforma disponibiliza o cadastramento de novas regras lógicas para notificações, como dito anteriormente, foi criada uma regra de inspeção que observa um relatório dinâmico de vacinas faltosas nas creches de Fortaleza, tornando possível comunicar via e-mail ao gestor cadastrado habilitado em receber essa notificação a relação das vacinas e quais são elas. Com posse dessas informações podem ser realizadas campanhas nas Unidades de Atenção Primária à Saúde (UAPS) mais próxima as creches.

Figura 6 - Quantidade de gestantes por bairros (Conjunto Ceará II)



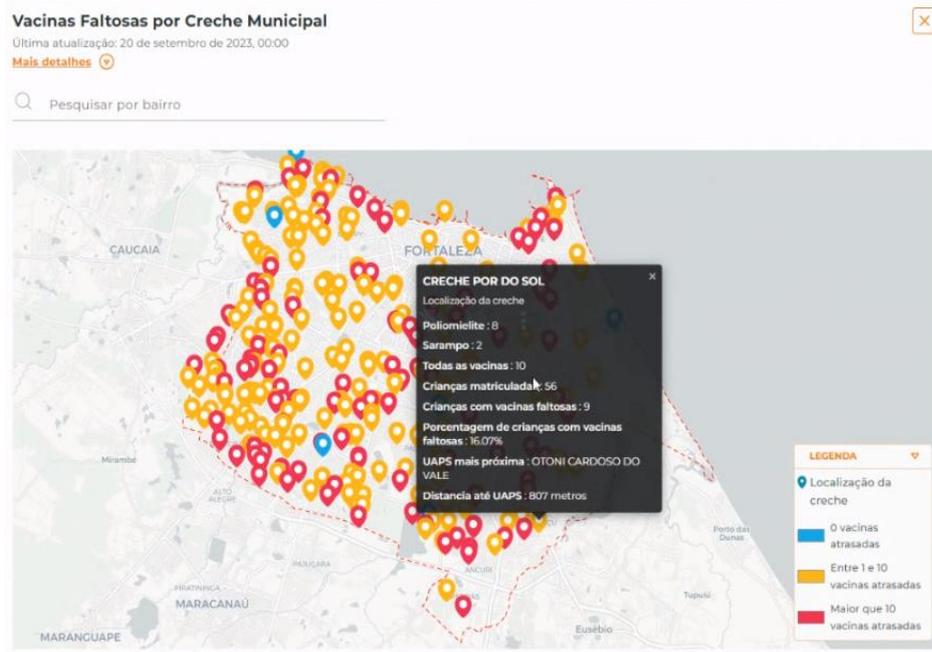
Fonte: BigData Fortaleza

Figura 7 - Quantidade de gestantes por bairros



Fonte: BigData Fortaleza

Figura 8 - Vacinas faltosas na creche Por do Sol.



Fonte: BigData Fortaleza

9 CONCLUSÃO

Os objetivos foram alcançados haja vista que se apresentou o uso de tecnologias como o *big data* para o desenvolvimento de cidades mais inteligentes. Se avaliou como a qualidade dos dados podem influenciar no processo de análise na hora da extração de evidências; sintetizou-se os métodos de análise de dados que podem ser utilizados na extração de evidências; se examinou os aspectos relacionados a privacidade e segurança no uso de *big data* na gestão pública; se explorou como o uso de *big data* pode facilitar e promover a participação cidadã no processo de interação entre governo e sociedade; se observou as oportunidades e desafios na transformação de cidades em ambientes mais inteligentes.

Constatou-se que nada mais insignificante que um único dado isolado, mas em grandes quantidades podem influenciar o processo decisório da administração pública em diversos campos, incluindo, economia, saúde, educação, moradia e mobilidade. Essa é a realidade nos dias atuais, nunca se produziu tantos dados, e esse fenômeno vem sendo impulsionado pelo uso massivo dos meios digitais, tais como, redes sociais, blogs, sistemas web e dispositivos de IoT.

Observou-se que os dados se bem tratado, podem ser a principal fonte de matéria-prima para tomada de decisões nas instituições, sejam elas públicas ou privadas, mas vimos que há uma longa jornada a ser explorada pelo setor público, principalmente, pois junto com esse crescimento novos desafios apareceram, e devem ser superados para a utilização de forma eficiente do *big data* nos processos decisórios, vire realidade.

De forma ampla, o conceito de *big data* pode ser definido como um conjunto tão grande e complexo de dados que se torna um desafio para sistemas tradicionais. Sendo assim, a administração pública terá que adaptar-se as novas tecnologias e técnicas de análise para aproveitar todo o potencial que esse volume de dados pode oferecer.

Notou-se que a cultura organizacional é um fator crítico a ser considerado. A adoção de tecnologias de *big data* muitas vezes requer mudanças de mentalidade e de processos organizacionais internos nas prefeituras, portanto, deve estar apta para não só apenas entender as novas tecnologias, como para moldar-se à nova realidade e adotar a cultura de dados, apesar da burocracia, da resistência à mudança e da falta de familiaridade com as ferramentas de *big data* representar obstáculos significativos para que essa mudança ocorra.

A pandemia do COVID-19, pegou o mundo de surpresa, e muitas prefeituras tiveram que se adaptar pelo aumento na demanda por serviços *on-line* pela população. E esse crescimento fomentou a busca de governos locais pela digitalização de serviços, buscando

solucionar a interação com a sociedade. Apesar de avanços alcançados pela digitalização, vários desafios ainda devem ser combatidos.

Viu-se que, a falta de padronização e integração dos dados existentes afetam diretamente a qualidade dos dados o que dificulta todo o processo em projetos de *big data*; falta cultura de dados nos servidores e colaboradores; o arcabouço de leis e regulamentos também é algo complexo e burocrático, podendo exigir um esforço adicional para garantir que os projetos estejam em conformidade legal; a necessidade de investimentos em tecnologia, principalmente a de infraestrutura de conectividade e cibersegurança; inexistência de um quadro técnico qualificado, são apenas alguns dos obstáculos a serem enfrentados pelas cidades.

Salienta-se que o principal desafio exposto nesse documento, é o apoio das lideranças locais, esse é um fator crítico para transformar a visão da administração pública, principalmente em um ambiente de escassez de recursos financeiros e mudanças políticas. Somente através da conscientização dessas lideranças será possível remover os obstáculos e superar os desafios, tornando possível a condução de projetos, como os de *big data*.

Outro aspecto necessário é garantir a institucionalização desses projetos, facilitando assim a sua continuidade. mesmo com mudança na gestão e tornando menos burocrático o acesso a recursos humanos e financeiros.

Foi exposto uma visão sobre como a aplicação efetiva das tecnologias voltadas ao aperfeiçoamento do acesso à informação de uma cidade devem ser plenas, visto que a internet se tornou um dos principais canais de informação e difusão de direitos sociais, sendo uma conquista do cidadão, sobretudo para os grupos mais vulneráveis.

Portanto, a gestão deverá promover a inclusão digital com amplo acesso aos serviços on-line, incluindo grupos desfavorecidos ou grupos que possam ter dificuldades no acesso à tecnologia, pois o analfabetismo digital é algo que aumenta as disparidades sociais e deve ser combatido pela sociedade brasileira.

Dentro desse contexto, políticas voltadas para a abertura de dados podem promover a inclusão digital e social da população. Essa política, garante à sociedade o acesso aos dados armazenados em bases públicas, promovendo a transparência, a participação popular e fortalece a legitimidade das políticas públicas.

No entanto, com o aumento da disponibilidade de serviços digitais, a administração pública deve proceder com cautela no manuseio de dados sensíveis da população, pois sempre devem se certificar de que os dados publicados não incluam informações pessoais que possam violar a privacidade dos cidadãos e que os dados sempre sejam usados de forma ética.

Analisou-se que as prefeituras muitas vezes têm recursos limitados, e a aquisição de tecnologia, bem como ter um quadro técnico qualificado, pode ser um ônus financeiro considerável. Mas se viu formas de minimizar esses custos, apesar de alguns aspectos como, a conectividade e o mínimo de pessoal qualificado serem críticos para soluções de *big data*. A formação de parceria entre governo-empresa ou governo-academia são formas de amenizar os custos e aumentar a capacidade técnica para o desenvolvimento desses projetos.

Viu-se também que a computação em nuvem pode representar uma alternativa viável e econômica ao município, principalmente em governos locais de pequeno porte, onde a aquisição e manutenção de uma infraestrutura de TIC própria, pode não ser uma alternativa viável em termo de custos. Adotar essa estratégia de priorização de serviços em nuvem podem reduzir os custos tecnológicos, além de garantir serviços 24x7 e segurança nos dados.

Considerando o momento, o grande volume de dados gerados, principalmente pela população; o aumento da possibilidade de coleta de dados em tempo real por meio de dispositivos IoT; as novas técnicas de análises, empregando técnicas de aprendizado de máquinas, que utilizam esses dados e são capazes de processá-los em curto tempo na tomada de decisões mais assertivas; o barateamento do custo de alguns serviços em nuvem, onde os provedores oferecem alguns deles de forma gratuita ou com carência de tempo para uso; linguagens *open source* como R e Python que são as mais utilizadas nessas análises; as parcerias que podem ser criadas. Ao integrar esses elementos ou parte deles, as cidades moldam o seu futuro de maneira mais sustentável e inteligente.

Em conclusão, a transformação das cidades em ambientes mais inteligentes, é algo inevitável, é só uma questão de tempo, algumas prefeituras mais rápidas do que outras. Isso foi o que motivou a criação dessa cartilha, tentando explicar de forma simples, mas esclarecedora a era em que se vive e como as prefeituras podem se beneficiar. Capacitando e incentivando os agentes locais a abraçarem essa tecnologia como uma ferramenta poderosa para impulsionar o desenvolvimento sustentável e a qualidade de vida em suas comunidades. Facilitando assim à transformação das cidades em ambientes mais inteligentes.

REFERÊNCIAS

- ABBASI, A.; SARKER, S.; CHIANG, R. H. L. **Big Data Research in Information Systems: Toward an Inclusive Research Agenda**. Journal of the Association of Information Systems, v. 17, n. 2, p. 1–32, 2016.
- ATKINSON, M.; PARSONS, M. Data-Intensive Thinking with DISPEL. In: **The Data Bonanza: Improving Knowledge Discovery in Science, Engineering, and Business**. M. ATKINSON, R.; BAXTER, M.; GALEA, M.; PARSONS, P.; BREZANY, O.; CORCHO, J.; HEMERT, V.; SNELLING, D. (Org.). USA: John Wiley & Sons, 2013.
- BETSER, Joseph; BELANGER, David. Architecting the enterprise via big data analytics. In: LIEBOWITZ, Jay. **Big data and business analytics**. Boca Raton: Taylor & Francis Group, LLC, 2013.
- BOUSKELA, Mauricio, et al. **Caminho para as Smart Cities**. Banco Interamericano de Desenvolvimento (2016).
- BUTTARELLI, Giovanni. **Meeting the challenges of big data: A call for transparency, user control, data protection by design and accountability**. European Data Protection. Brussels: EDPS - European Data Protection Supervisor, 2015. 21 p. Disponível em: <https://edps.europa.eu/sites/edp/files/publication/15-11-19_big_data_en.pdf>. Acesso em: 24 jul. 2023.
- CERDEIRA, Pablo de Camargo; MENDONÇA, Marcus Mentzingen de; LAGOWSKA, Urszula Gabriela. **Políticas públicas orientadas por dados: os caminhos possíveis para governos locais**. BID, 2020.
- CHEN, D. Q.; PRESTON, D. S.; SWINK, M. **How the Use of Big Data Analytics Affects Value Creation in Supply Chain Management**. Journal of Management Information Systems, v. 32, n. 4, p. 4–39, 2015.
- DAVENPORT, H. **Big data no trabalho: derrubando mitos e escolhendo oportunidades**. Rio de Janeiro: Elsevier, 2014.
- DATA SCIENCE ACADEMY. **Deep Learning Book** Disponível em: <https://www.deeplearningbook.com.br/o-que-sao-redes-neurais-artificiais-profundas/>. Acesso em: 07 set. 2023.
- DE LUCA, Cristina; BASSI, Silvia. **Potencializando o uso de big data para cidades inteligentes: um guia estratégico para gestores**. 2023.
- DINIZ. **O governo eletrônico no Brasil: perspectiva histórica a partir de um modelo estruturado de análise**. Revista de Administração Pública-RAP, Rio de Janeiro, n. 43, v. 1, p. 23-48, jan./fev. 2009. Disponível em: http://www.scielo.br/scielo.php?pid=S0034-76122009000100003&script=sci_abstract&tlng=pt. Acesso em: 05 jun. 2023.
- DOS DEPUTADOS, Câmara. **Cidades Inteligentes: Uma abordagem humana e sustentável**. Edições Câmara, 2021. Disponível em:

[https://www.andusbrasil.org.br/acervo/publicacoes/67-cidades-inteligentes-uma-abordagem-humana-e-sustentavel-camara-dos-deputados#:~:text=Acervo-,%22Cidades%20Inteligentes%3A%20Uma%20Abordagem%20Humana%20e%20Sustent%3A1vel%22%20\(C%3%A2mara,C%3%A2mara%20dos%20Deputados%20\(Cedes\)](https://www.andusbrasil.org.br/acervo/publicacoes/67-cidades-inteligentes-uma-abordagem-humana-e-sustentavel-camara-dos-deputados#:~:text=Acervo-,%22Cidades%20Inteligentes%3A%20Uma%20Abordagem%20Humana%20e%20Sustent%3A1vel%22%20(C%3%A2mara,C%3%A2mara%20dos%20Deputados%20(Cedes).). Acesso em: 12 ago. 2023.

ELMASRI, R.; NAVATHE, S. B. **Sistemas de Banco de Dados**. São Paulo: Pearson, 2011.

GOMES; Gabriela Tavares; FERTIG, Max Romoaldo. **Construção de uma base de conhecimento de dados governamentais abertos baseada em ontologia utilizando dados conectados**. Trabalho de Conclusão de Curso (Graduação em Sistemas de Informação), Universidade do Sul de Santa Catarina, Florianópolis, 2016.

GARTNER. **Big data**. Disponível em: <<https://www.gartner.com/en/information-technology/glossary/big-data>>. Acesso em: 07 jul. 2023.

GARTNER **Quadrante mágico para infraestrutura em nuvem e serviços de plataforma** Disponível em: <<https://www.gartner.com/technology/media-products/reprints/AWS/1-2AOZQARL-PTB.html?refid=f70460f2-4ce1-473b-bf28-3e279852fb66>>. Acesso em: 07 jul. 2023.

GROVER, V.; CHIANG, R. H. L.; LIANG, T.; ZHANG, D. **Creating Strategic Business Value from Big Data Analytics: a research Framework**. Journal of Management Information Systems, v. 35, n. 2, p. 388-423, 2018.

HAMMOND, K. **The Value of Big Data Isn't the Data**. Blog da Harvard Business. Review, May 2013. Disponível em: <http://blogs.hbr.org/cs/2013/05/the_value_of_big_data_isnt_the.html>. Acesso em: 5 jun. 2023.

HERNANDEZ, M. J. **Database Design for Mere Mortals: a hands-on guide to relational database design**. 25th edition. Boston: Pearson, 2021.

International Telecommunication Union. 2016. **Focus Group on Smart Sustainable Cities**. [online] Disponível em: <http://www.itu.int/en/ITU-T/focusgroups/ssc/Pages/default.aspx>.

KIM, Gang-Hoon, Silvana TRIMI, and Ji-Hyong CHUNG. **Big-data applications in the government sector**. Disponível em: <<https://linkedkey.com/wp-content/uploads/2016/05/Big-Data-Applications-in-the-Government-Sector.pdf>>. Acesso em: 18 jul. 2023.

MACHADO, F. N. R. **Big Data: o Futuro dos Dados e Aplicações**, 1. ed. São Paulo: Saraiva, 2018

MILANI, C. R. S. **O princípio da participação social na gestão de políticas públicas locais: uma análise de experiências latino-americanas e europeias**. Revista de Administração Pública, v. 42, n. 3, p. 551–579, jun. 2008.

MINELI, Michael; CHAMBERS, Michele; DHIRAJ Ambiga. **Big data, big analytics: emerging business intelligence and analytic trends for today's businesses**. New Jersey: John

Wiley & Sons, Inc., 2013.

OPEN KNOWLEDGE INTERNATIONAL. Disponível em: <<https://ok.org.br/dados-abertos>>. Acesso em: 07 jul. 2023.

PEREIRA, Gabriela Viale. **Contribuição de iniciativas de cidades inteligentes no desenvolvimento humano: uma análise da percepção de agentes de centros de operações municipais no Brasil**. 2016. Tese (Doutorado em Administração) –Faculdade de Administração, Contabilidade e Economia, Pontifícia Universidade Católica do Rio Grande do Sul, Rio Grande do Sul, 2016. Disponível em: <http://repositorio.pucrs.br/dspace/bitstream/10923/8307/1/000478655-Texto%2bCompleto-0.pdf>. Acesso em: 12 ago. 2023.

SCHÖNBERGER-MAYER, Viktor; CUKIER Kenneth. Tradução Paulo Palzonoff Junior. **Big data: como extrair volume, variedade, velocidade e valor da avalanche de informação cotidiana**. Rio de Janeiro: Elsevier, 2013.

TAURION, Cezar. **Big data**. Rio de Janeiro: Brasport, 2013.

TRIBUNAL DE CONTAS DA UNIÃO – TCU. Secretaria de Fiscalização de Tecnologia da Informação. **5 motivos para a abertura de dados na Administração Pública**. Brasília, 2015. Disponível em: <<http://portal3.tcu.gov.br/portal/pls/portal/docs/2689107.PDF>>. Acesso em: 11 jul. 2023.

ZUIDERWIJK, Anneke; JANSSEN, Marijn. **Open data policies, their implementation and impact: A framework for comparison**. *Government Information Quarterly*, 31(1), pp. 17-29, 2014. Disponível em: <<https://www.sciencedirect.com/science/article/abs/pii/S0740624X13001202?via%3Dihub>>. Acesso em: 07 jul. 2023.

APÊNDICE A – BIG DATA: OPORTUNIDADES E DESAFIOS PARA GOVERNOS LOCAIS (CARTILHA)





UNIVERSIDADE FEDERAL DO CEARÁ
PROGRAMA DE PÓS-GRADUAÇÃO EM ECONOMIA (PROFISSIONAL)
MESTRADO PROFISSIONAL EM ECONOMIA DO SETOR PÚBLICO - MESP

BIG DATA: OPORTUNIDADES E DESAFIOS PARA GOVERNOS LOCAIS

Essa cartilha faz parte da dissertação apresentada ao Programa de Pós-Graduação em Economia – CAEN, Curso de Mestrado Profissional em Economia, da Universidade Federal do Ceará – UFC, como requisito parcial para obtenção do título de Mestre em Economia. Área de Concentração: Economia do Setor Público.

Prof. Dr. Fabrício Carneiro Linhares (Orientador)
Universidade Federal do Ceará (UFC)

Joaquim Nerson Moura Filho (Orientando)
Universidade Federal do Ceará (UFC)

SUMÁRIO

- 04 *BIG DATA: Por que deveria me importar?*
- 08 *Por que um gestor público de uma cidade pequena brasileira deveria usar?*
- 11 *Dados de qualidade.*
- 12 *Política de dados abertos.*
- 15 *Análises de dados.*
- 19 *Explicando o processo de aprendizado de máquina.*
- 23 *Cidade inteligente.*
- 25 *Desafios.*
- 28 *Falando um pouco mais de tecnologia.*
- 32 *Exemplo prático.*
- 36 *Conclusão.*
- 37 *Referências bibliográficas.*

BIG DATA:

POR QUE DEVERIA ME IMPORTAR?



APRESENTAÇÃO.

Poucas transformações trazem mudanças tão significativas para a sociedade como os avanços tecnológicos que ultrapassam as barreiras da inovação, modificando profundamente a forma de interação entre as organizações e as pessoas.

O mundo viveu quatro transformações nesse sentido, chamadas Revoluções Industriais que provocaram mudanças profundas na sociedade, moldando a forma que vivemos, trabalhamos e nos relacionamos nas diversas dimensões, incluindo econômica, social, cultural e política.

A era digital começou na Terceira Revolução Industrial.

Hoje, estamos na quarta revolução, que adota tecnologias avançadas, como, internet das coisas, inteligência artificial, big data, dentre outras.

E sua principal matéria-prima são os dados.

A ERA DOS DADOS.

Estima-se que desde o início da civilização até 2003 foram produzidos pela humanidade 5 exabytes de dados. Mas segundo o International Data Corporation (IDC), criamos atualmente esse mesmo volume de dados a cada dois dias e espera-se que a quantidade de dados gerados no mundo alcance 175 zettabytes até 2025.

Vivemos na era dos dados e das informações em potencial, impulsionado pelo uso massivo dos meios digitais, tais como, redes sociais, blogs, sistemas web e dispositivos IoT (internet das coisas).



TEMOS:



Um grande
volume
de dados



Uma grande
variedade
de dados



Produzidos
em grande
velocidade

QUE PRECISAM TER
 **veracidade**

PARA EXTRAIR
 **valor**

ISSO É BIG DATA.

O conceito de big data pode ser definido como um conjunto tão grande e complexo de dados que se torna um desafio para sistemas tradicionais de processá-los, armazená-los e analisá-los de forma eficaz e produtiva.

O termo “Vs do big data” refere-se a características-chave que são frequentemente usadas para descrever os desafios e as propriedades distintas do armazenamento, processamento e análise de grandes conjuntos de dados.



ALÉM DE QUE...

Big data não é uma tecnologia específica, mas um conjunto de tecnologias em contínuo desenvolvimento ao longo do tempo, que agora são suportadas por infraestruturas capazes de lidar com a vasta quantidade e complexidade dos dados gerados em ambientes distribuídos e não convencionais, e a **computação distribuída** desempenhou um papel fundamental nesse processo.

E QUAIS SÃO ESSAS TECNOLOGIAS?

APACHE HADOOP

O poder do processamento paralelo e distribuído



Apache Hadoop, um framework para o processamento e armazenamento de grandes quantidades de dados em clusters de forma ágil, escalável e tolerante a falhas, garantindo a integridade dos dados mesmo em casos de indisponibilidades de clusters e tudo isso de maneira transparente.

O processamento e armazenamento de forma paralela e distribuída é o grande diferencial desse framework e foi a principal solução que viabilizou o custo do uso de big data.

Benefícios:

- Escalável;
- Flexível;
- Open-Source;
- Tolerante a falhas.

BANCOS NoSQL



O armazenamento tradicional, que trata dados estruturados, geralmente usados pelos bancos de dados relacionais não são mais capazes de absorver dados gerados por fontes como redes sociais e sensores, por exemplo, que em sua grande maioria criam dados não estruturados ou semiestruturados.

Para solucionar esse problema surgiu como um paradigma não-tradicional os bancos de dados NoSQL (Not Only SQL), capazes de lidar com os grandes volumes e variedade de dados e resolver alguns dos desafios que surgiram para implementar projetos de big data.

E QUAIS SÃO ESSAS TECNOLOGIAS?

DISPOSITIVOS IoT



A **internet das coisas (IoT)** é a capacidade de conectar à internet sensores usados no dia a dia das pessoas e organizações, podendo coletar e transmitir dados, criando um ambiente mais inteligente.

Com a crescente adoção de dispositivos IoT a quantidade de dados coletados em tempo real aumentou significativamente, tornando-se um aspecto crucial para análises de big data, permitindo a detecção de padrões, tendências e insights instantâneos.

A imensa rede criada por dispositivos IoT tem mudado como tratamos informações e interações atualmente, trazendo inovação e valor agregado aos processos e operações do dia a dia.

5G



Com o avanço da quantidade de dispositivos conectados à internet, surge a necessidade de um aprimoramento das redes móveis, a tecnologia 5G é a próxima geração de redes de comunicação móvel que oferece velocidades muito mais rápidas, com menor latência e maior capacidade de conexão em comparação com as gerações anteriores, sendo fundamental para lidar com o volume de dados gerados pelos dispositivos IoT.

O 5G utiliza frequências de rádio mais altas, tecnologias avançadas de modulação e antenas mais sofisticadas para proporcionar taxas de transferência de dados muito mais rápidas em comparação com as gerações anteriores.



POR QUE UM GESTOR PÚBLICO DE UMA PEQUENA CIDADE BRASILEIRA DEVERIA USAR?

- Tomada de decisões baseadas em evidências = + **Dados – Achismos.**
- Melhorando os serviços e políticas públicas prestados à população.
- Economia de recursos.
- Eficiência operacional, devido ao compartilhamento de dados entre as secretarias.
- Aumenta a transparência.
- Maior engajamento da população.
- Aumenta da confiança na administração pública.

As administrações públicas possuem acesso a uma vasta quantidade de dados, próprias ou de terceiros, trazendo uma oportunidade sem igual, que embasam os dados na tomada de decisão, desempenhando assim um papel essencial no aprimoramento de políticas públicas.

**DESCUBRA SEUS DADOS
DESCONHECIDOS E GERE
EVIDÊNCIAS COM ELES.**

**E NAS MAIS
VARIADAS ÁREAS
DE UMA CIDADE,
RESOLVENDO
MUITOS DOS
DESAFIOS COMUNS.**



NO BRASIL:



Havia 181,8 milhões de usuários de Internet no Brasil no início de 2023, quando a penetração da Internet era de 84,3%.

Tínhamos 152,4 milhões de usuários de mídias sociais em janeiro de 2023, o que equivale a 70,6% da população total.

Um total de 221,0 milhões de conexões móveis celulares estavam ativas no Brasil no início de 2023, valor equivalente a 102,4% da população total.

Fonte: Digital 2023 Global Overview Report
<https://datareportal.com/reports/digital-2023-global-overview-report>

O GESTOR QUE SUBESTIMA A CAPACIDADE DE IDENTIFICAR OS DESEJOS E ANSEIOS DA POPULAÇÃO POR MEIO DOS DADOS GERADOS POR ELA MESMA ESTÁ PERDENDO UMA VALIOSA OPORTUNIDADE.

ALGUMAS INICIATIVAS JÁ FORAM TOMADAS.

Uma colagem de várias imagens de notícias e sites brasileiros que destacam o uso de Big Data e Inteligência Artificial em diferentes setores. As imagens incluem:

- Ministério da Justiça entrega plataforma de Big Data e Inteligência Artificial desenvolvida pela UFC.
- FGV PROJETOS: Uso de Big Data ajuda governo brasileiro a gastar de forma mais eficiente.
- UNIVERSIDADE FEDERAL DO CEARÁ: Plataforma BigData para direção e fiscalização pela Prefeitura.
- Cidade gaúcha cria laboratório de data science para orientar políticas públicas.
- Moovit apresenta panorama em big data do transporte público em 10 grandes cidades brasileiras.
- Big data auxilia o desenvolvimento de políticas públicas (exame.com).
- Como o uso de Big Data pode ajudar a identificar padrões em doenças.
- Medicinas: SUS terá modelo de dados para registrar procedimentos, consultas e cirurgias.

OK, BASTA TER DADOS ENTÃO? INFELIZMENTE NÃO. DADOS NÃO SÃO INFORMAÇÃO, MAS PODEM SE TORNAR. ENTÃO, TEREMOS O VALOR DO BIG DATA.



MINHA CIDADE É PEQUENA, SERÁ QUE CONSIGO REALIZAR UM PROJETO USANDO BIG DATA?

Claro que consegue. Inicie pequeno, mas pensando grande.

APOIO DAS LIDERANÇAS LOCAIS



APOIO DA COMUNIDADE

E POR ONDE COMEÇO? DEFINA OS PROBLEMAS.

Identifique os desafios que a gestão pública enfrenta, priorize com base na sua importância e impacto na comunidade. Desenvolver critérios claros para priorizar os desafios é fundamental, isso pode incluir a gravidade do problema, o impacto na qualidade de vida dos cidadãos, a viabilidade de soluções, o custo de implementação e a urgência.

Os desafios atuais são diferentes daqueles do passado e as perspectivas de solucioná-los também são outras, **procure nos dados.**

A técnica dos “5 Porquês”, pode ser útil, essa é uma abordagem de resolução de problemas que busca identificar a causa raiz de um problema ou desafio.

DADOS DE QUALIDADE.

Após os levantamentos iniciais, é hora de listar os dados que serão necessários para o projeto, isso pode incluir dados governamentais, dados de cidadãos, dados de sensores, entre outros. Obtenha acesso em dados pertencentes a terceiros e verifique as recomendações legais necessárias para obter autorização de acesso.

Como atualmente a maioria dos dados gerados são do tipo não estruturado, que possuem como característica a heterogeneidade de formatos, a simples coleta não é mais o bastante. As organizações devem ter a preocupação com a qualidade desses dados coletados.

O valor como uma das características-chave do big data está ligado diretamente a qualidade dos dados.

De acordo com Kim, Trimi e Chung (2014), com um vasto volume de dados armazenados em seus bancos de dados, o governo possui a oportunidade de analisá-los minuciosamente e extrair insights que podem resultar na criação de serviços inovadores e no aprimoramento de sua administração.

PARA QUE ELES ATENDAM A ESSE REQUISITO ALGUMAS CARACTERÍSTICAS SÃO NECESSÁRIAS.

a) precisão: o dado deve ser exato, confiável, devendo ser livres de erros;

b) disponível no tempo: devem refletir a realidade do momento, para não levar a decisões baseadas em dados obsoletas;

c) relevante: os dados devem refletir a realidade do contexto em que serão usados;

d) completo: os dados não devem ser corrompidos ou incompletos de forma a prejudicar a sua utilidade;

e) simples: os dados devem ser apresentados da forma mais clara e compreensível;

f) confiável: os dados não devem mudar de forma inesperada ou inconsistente.

VOCÊ CONHECE A POLÍTICA DE DADOS ABERTOS?

A política de dados abertos visa garantir e facilitar à sociedade o acesso aos dados produzidas ou custodiadas pelos órgãos públicos, com o intuito de promover a transparência, a participação popular e até mesmo o desenvolvimento de serviços ou aplicações pelo uso ou reuso desses dados.

A abertura de dados públicos não aumenta apenas a transparência, mas também a participação popular, ao promover a circulação dos dados entre as organizações públicas e a sociedade, cria-se assim um ambiente de reuso desses dados.

Permitir que os dados sejam reutilizáveis cria um ambiente onde novos produtos ou serviços possam ser desenvolvidos pela sociedade, impulsionando assim um ciclo de inovação. Dessa forma as organizações públicas podem se beneficiar com os dados coproduzidos pela sociedade para elaboração de políticas ou serviços públicos.

Quando essa ampla variedade de dados do governo é aberta, fomenta a participação popular. Segundo o Tribunal de Contas da União (2015, p.13), assim organizações, cidadãos, acadêmicos e até mesmo instituições públicas têm a possibilidade de utilizar bases de dados públicos para a produção e o compartilhamento de novos conhecimentos e de novos serviços, numa concepção de coparticipação entre ente privado e governo na oferta de serviços públicos à sociedade.

POLÍTICA DE DADOS ABERTOS NÃO É A MESMA COISA DE PORTAIS DA TRANSPARÊNCIA?

Apesar de apresentar características relacionados, contribuindo para o aumento da transparência e o acesso às informações governamentais, portais da transparência e políticas de dados abertos possuem focos diferentes.

Portais da transparência têm como foco disponibilizar informações sobre as atividades financeiras e administrativas do governo, isso inclui, por exemplo, contratos, licitações, salários de servidores, permitindo que a sociedade faça o acompanhamento desses gastos.

Já as políticas de dados abertos têm um foco mais amplo, visando promover a abertura de dados governamentais de maneira geral, não se limitando apenas com informações financeiras e administrativas.

EXISTE ALGUNS PRINCÍPIOS FUNDAMENTAIS PARA OS DADOS ABERTOS GOVERNAMENTAIS.

- a) completos:** todos os dados públicos estão disponíveis, ou seja, dado público é aquele que não está sujeito a limitações válidas de privacidade, segurança ou controle de acesso;
- b) primários:** os dados são apresentados tais como os coletados na fonte, com o maior nível de granularidade e sem agregação ou modificação;
- c) atuais:** são disponibilizados tão rapidamente quanto necessária à preservação do seu valor;
- d) acessíveis:** são disponibilizados para o maior alcance possível de usuários e para o maior conjunto possível de finalidades;
- e) compreensíveis por máquinas:** são razoavelmente estruturados de modo a possibilitar processamento automatizado;
- f) não discriminatórios:** são disponíveis para todos, sem exigência de requerimento ou cadastro;
- g) não proprietários:** são disponíveis em formato sobre o qual nenhuma entidade detenha controle exclusivo;
- h) livres de licenças:** não estão sujeitos a nenhuma restrição de direito autoral, patente, propriedade intelectual ou segredo industrial, mas as restrições sensatas relacionadas à privacidade, segurança e privilégios de acesso devem ser permitidas.

EXEMPLOS DE DADOS ABERTOS:

GOVERNO FEDERAL

<https://dados.gov.br/home>



IBGE

<https://sidra.ibge.gov.br/>



BANCO CENTRAL

<https://dadosabertos.bcb.gov.br/>



IPEA

<http://www.ipeadata.gov.br/Default.aspx>



SSP SÃO PAULO

<https://www.ssp.sp.gov.br/transparenciassp/>



LEMBRE-SE DE QUE...

A utilização de dados abertos capacita a administração pública na utilização do big data, ao promover credibilidade regulatória do acesso e uso.

Como representa mudanças significativas no ambiente organizacional da gestão e na própria relação entre governo e sociedade, para que projetos baseados em dados tenham sucesso, é necessário que seja criada com antecedência uma cultura governamental capaz de assimilar esse novo paradigma.

Acabe com os silos de dados, quando os dados estão desacoplados e espalhados em várias secretarias, cria-se uma grande dificuldade em reunir todas as informações necessárias para se obter uma análise eficiente em uma solução inteligente baseada em dados.

Uma questão que deve ser observada nas políticas de dados abertos é a garantia que o acesso aos dados seja amplo, incluindo também, grupos desfavorecidos ou grupos que possam ter dificuldades no acesso à tecnologia, sendo necessário a inclusão social e digital desses grupos por parte do setor público.

No entanto, esse potencial só será totalmente explorado se os dados públicos forem verdadeiramente abertos, sem restrições, legais, financeiras ou tecnológicas, para garantir a inclusão de todos na reutilização desses dados.

E O MAIS IMPORTANTE, GARANTA QUE OS DADOS SEJAM COLETADOS E USADOS DE FORMA ÉTICA E ESTEJAM EM CONFORMIDADE COM AS REGULAMENTAÇÕES DE PRIVACIDADE.

“essa transformação não só dificulta a proteção da privacidade como também apresenta uma ameaça totalmente nova: castigos com base em propensões, isto é, a possibilidade de usar previsões de big data sobre pessoas para julgá-las e puni-las antes mesmo que elas ajam, o que renega a ideia de justiça e livre-arbítrio. Além da privacidade e da propensão, há um terceiro perigo. Corremos o risco de sermos vítimas da ditadura dos dados, na qual adoramos as informações e os resultados de nossas análises e acabamos usando-os de forma equivocada. Com responsabilidade, o big data é um instrumento útil de tomada de decisão. Se usados sem sabedoria, eles se tornam um instrumento de poderosos, que podem transformá-los numa fonte de repressão, seja ao frustrar clientes e funcionários ou, pior, ao atacar cidadãos.” Schönberger-Mayer e Cukier (2013, p. 105 - 106).

QUEM NÃO TEM CONHECIMENTO DOS DADOS QUE POSSUI, SERÁ INCAPAZ DE GARANTIR SUA SEGURANÇA.

E AGORA, O QUE FAÇO COM OS DADOS?

Chegou a hora de criar valor para os seus dados. A hora de usar evidências nas decisões. Chegou a hora da análise de dados.

Mas lembre-se, que a capacidade de ter feito perguntas relevantes para definir quais problemas serão resolvidos é o passo fundamental para determinar quais ferramentas, técnicas e modelos preditivos deverão ser utilizados de forma adequada para um processo bem-sucedido de análise de dados.

OS 4 TIPOS DE ANÁLISES DE DADOS.



NÃO É NECESSÁRIO CRIAR ALGO EXTRAORDINÁRIO, CRIE APENAS UM MODELO QUE FAÇA UMA ATIVIDADE MELHOR DO QUE ATUALMENTE ESTEJA SENDO FEITA, E LEMBRE-SE QUE O BÁSICO PODERÁ SER O SUFICIENTE.

E COMO EU PREVEJO O FUTURO?

Agora é o momento de aproveitar o aprendizado de máquinas, mas conhecido como **Machine Learning**.

É um subconjunto da inteligência artificial que tenta aprender através do treinamento de modelos usando grandes conjuntos de dados (big data) para que possam aprender a reconhecer e generalizar padrões, com o objetivo de tomar decisões ou fazer previsões precisas ao utilizar novos dados.

E o segredo é simples, a projeção de um resultado no futuro se dá a partir da investigação de dados passados, o sucesso dessas análises se deve ao big data, pois quanto mais dados forem oferecidas para essas análises mais precisas serão as previsões.

Mas fica a dica. Como elas são baseadas em probabilidade nunca serão completamente precisas, ou seja, elas preveem o que pode acontecer no futuro com um nível aceitável de confiabilidade que inclui cenários hipotéticos e avaliação de riscos.

QUAIS OS TIPOS DE APRENDIZADO DE MÁQUINA?

APRENDIZADO
SUPERVISIONADO

APRENDIZADO
NÃO SUPERVISIONADO

APRENDIZADO
POR REFORÇO

APRENDIZADO SUPERVISIONADO.



Ocorre quando o algoritmo aprende a partir de dados de exemplos com inputs e possíveis outputs que podem conter valores quantitativos ou qualitativos afim de prever a resposta correta quando recebe novos dados.

O algoritmo recebe um conjunto de entradas, juntamente com as saídas corretas correspondentes e o algoritmo aprende comparando a sua saída real com as saídas corretas. Em seguida, o algoritmo ajusta o modelo de acordo com seu processo de aprendizagem.

Esse tipo de aprendizagem pode ser subdividido em dois tipos, **os algoritmos de classificação** são usados para o treinamento de modelos onde a saída é uma classe ou grupo à qual um dado pertence; e os **algoritmos de regressão** onde a saída do modelo é um valor numérico que caracteriza a relação entre as variáveis independentes e a variável alvo.

APRENDIZADO NÃO SUPERVISIONADO.

Ocorre quando o algoritmo aprende com exemplos simples, mas sem qualquer resposta associada, deixando a cargo do algoritmo determinar os padrões de dados por conta própria, este tipo de algoritmo tende reestruturar os dados com novos recursos que podem representar uma nova classe ou uma série de valores não correlacionados, são muito úteis a fornecer insights sobre o significado dos dados, ou seja, quando o algoritmo pode automaticamente encontrar relações ou padrões em um conjunto de dados.

Em vez de prever uma variável alvo, como no aprendizado supervisionado, o aprendizado não supervisionado busca identificar relações naturais entre as amostras de dados com o objetivo de organizar os dados de alguma forma ou descrever sua estrutura.

Existem **algoritmos de associação** onde são capazes de identificar padrões entre os dados, como por exemplo, criar uma associação entre produtos que geralmente são comprados juntos, normalmente são utilizadas para explicar padrões entre informações independentes. Já os **algoritmos de clusterização** são usados para descobrir grupos ocultos nos dados, categorizando as informações a fim de gerar segmentações e com isso agrupá-las por características parecidas.

APRENDIZADO POR REFORÇO.



É um aprendizado por interação, similar ao que chamamos de aprender por tentativa e erro, é um paradigma avançado no campo da aprendizagem de máquina que se concentra em como agentes autônomos podem aprender a tomar decisões sequenciais para otimizar sua interação com um ambiente, treinando para tomar decisões em busca dos melhores resultados

Os componentes envolvidos nesse tipo de aprendizagem são, o agente (tomador de decisão), o ambiente (onde ocorre a interação com o agente) e as ações (que são as decisões que o agente pode escolher para que maximizar a premiação esperada). Tendo como objetivo fazer com que o agente escolha opções que maximizem a recompensa esperada ao longo de um determinado período de tempo.

A aprendizagem por reforço é frequentemente usada em cenários em que um agente deve tomar uma série de decisões sequenciais, como em jogos, robótica, controle de sistemas autônomos (carros autônomos).

APRENDIZADO PROFUNDO (DEEP LEARNING).



Aprendizagem profunda emerge como um subdomínio da aprendizagem de máquina envolvendo redes profundas, destacando-se pela capacidade de aprender e extrair informações complexas e hierárquicas a partir de grandes volumes de dados e reconhecer padrões que podem ser difíceis ou impossíveis de serem identificados pelo ser humano.

É uma das áreas de alto impacto da aprendizagem onde se tenta imitar o funcionamento do cérebro humano no processamento de dados e na criação de padrões para uso na tomada de decisões envolvendo uma arquitetura de redes neurais artificiais que são capazes de aprender sem supervisão humana.

Ela se concentra em treinar algoritmos conhecidos como redes neurais artificiais (RNAs) para aprender e representar automaticamente os dados em múltiplas camadas de abstração, sendo eficaz em tarefas complexas que envolvam grandes volumes de dados. Essas redes possuem camadas intermediárias, denominadas camadas ocultas, que desempenham um papel crucial na extração de características complexas dos dados.

EXPLICANDO O PROCESSO DE APRENDIZADO DE MÁQUINA.

Defina qual pergunta você deseja responder, com essa informação saberemos quais dados serão coletados, quais ferramentas, técnicas e modelos preditivos deveram ser utilizados de forma adequada para um processo bem-sucedido de análise de dados.

Antes de começar a coletar dados ou escolher algoritmos, é fundamental ter uma compreensão clara do que se espera alcançar com o modelo de machine learning.

Quanto mais específica a pergunta, mais fácil será medir o sucesso do modelo. Evite questões amplas e indefinidas.

Qual é o problema específico que você está tentando resolver?

Como o sucesso será medido?

Vale lembrar que a pergunta será feita para qualquer tipo de análise e não apenas para as preditivas.

COLETE OS DADOS NECESSÁRIOS.



Essa é uma etapa essencial, pois devemos garantir a qualidade dos dados e geralmente esses dados necessitam passar por um processo chamado ETL (Extração, Transformação e Carga), esse é um processo que garante a integridade e a padronização dos dados, facilitando as análises.

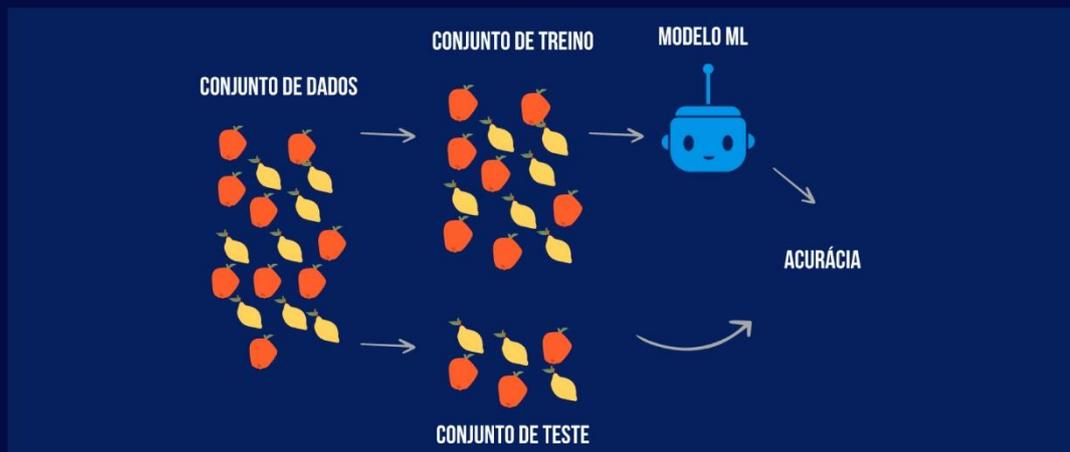
Extração processo de extrair os dados relevantes nas fontes de dados de origem.

Transformação é um processo de limpeza, correção, padronização dos dados para que os mesmos sejam mais efetivos nas análises, é um processo demorado.

Carga é o carregamento dos dados transformados em um banco de dados que será usado no processo de análise.

O ETL é opcional, podemos carregar os dados de uma fonte confiável que já tenha realizado as transformações necessárias.

AGORA VAMOS DIVIDIR A TOTALIDADE DOS DADOS HISTÓRICOS GERADOS NA CARGA.



Dividimos o conjunto de dados em dois conjuntos menores, um conjunto de treino e um outro conjunto de teste.

E por que isso?

Algoritmos de machine learning aprendem com dados, sendo que a principal razão, é avaliar como o modelo generaliza para dados que não foram usados durante o treinamento, ou seja, é realizar previsões precisas em dados que nunca viu antes.

Se usarmos todos os dados para treinar o modelo, ele pode se ajustar demais a esses dados e não conseguirá generalizar bem para um novo conjunto de dados.

Dados de treino são os dados que serão utilizados para treinar o algoritmo de machine learning para gerar um modelo preditivo. Geralmente são utilizados 70% da totalidade dos dados.

Dados de teste serão apresentados ao modelo após a sua criação, simulando previsões, já que são um conjunto de dados diferentes do conjunto de treino, permitindo verificar o seu desempenho. Geralmente são utilizados 30% da totalidade dos dados.

IDENTIFICANDO O MELHOR ALGORITMO PARA O APRENDIZADO DE MÁQUINA.

A escolha do algoritmo de aprendizado de máquina depende fortemente da natureza específica do problema que você está tentando resolver.

Alimente o conjunto de treinamento no modelo, permitindo que ele ajuste seus parâmetros para aprender padrões.

Avalie o desempenho do modelo usando o conjunto de teste.

Existem várias técnicas e métricas para avaliar o desempenho de um modelo de machine learning (acurácia, precisão, matriz de confusão, recall, F1-score para classificação; erro médio absoluto, erro quadrático médio para regressão).

Implemente o modelo.

Algoritmos não são modelos, por exemplo, caso você estivesse usando um algoritmo de aprendizado supervisionado, como a regressão linear, o algoritmo seria a fórmula matemática usada para ajustar uma linha aos dados, enquanto o modelo seria a linha específica ajustada aos seus dados após o treinamento.

RESUMINDO, O QUE AS ANÁLISES PODERIAM AJUDAR A RESPONDER?

Existem padrões incomuns nos gastos públicos?

Podemos detectar anomalias que possam indicar atividades não usuais ou possíveis irregularidades?

Existem indícios de sonegação no pagamento de tributos?

Como as tendências nas taxas de criminalidade evoluem ao longo do tempo?

Podemos antecipar picos e alocar recursos adequadamente?

Se aumentarmos o número de policiais em um determinado bairro, isso resultará em uma redução na taxa de criminalidade?

Dado o desempenho passado dos alunos em testes padronizados, podemos criar um modelo para prever quais alunos estão em risco de não atingir as metas de aprendizado?

OBSERVAÇÕES IMPORTANTES.

Embora a tomada de decisão baseada em dados seja geralmente considerada uma abordagem confiável e informada, ela não deve ser encarada como a única fonte de informação.

Geralmente quando usados algoritmos de aprendizado de máquinas, as entradas são conhecidas e em alguns casos os resultados, mas como eles aprendem padrões a partir de dados sem implementação explícita para isso, não se sabe com exatidão quais regras são usadas por eles, podendo preservar vieses.

Eles também não podem definir valores, objetivos, prioridades e as metas específicas da organização ou da sociedade sendo necessário embasar decisões alinhadas com esses valores.

E sempre deve ser considerado os aspectos éticos envolvidos, sendo fundamental garantir a privacidade, a segurança e a equidade no uso dos dados, além disso, é necessário considerar questões de justiça social e evitar o viés nos dados e nos algoritmos utilizados.

FAZENDO ISSO EU TENHO UMA CIDADE INTELIGENTE?

Já seria um começo bastante promissor, apesar de não só ser apenas isso, mas você já teria uma cidade mais inteligente.

Uma Cidade Inteligente e sustentável é uma cidade inovadora que utiliza as Tecnologias de Informação e Comunicação (TIC) e outros meios para melhorar a qualidade de vida, a eficiência das operações e serviços urbanos e sua competitividade, enquanto garante o atendimento das necessidades das gerações atuais e futuras com relação aos aspectos econômicos, sociais e ambientais. Fonte: International Telecommunication Union (Focus Groups on Smart Sustainable Cities, 2014)

São cidades capazes de colocar a população no centro do desenvolvimento, incorporando tecnologias na gestão urbana e utilizando esses elementos na estruturação de um governo eficiente, que engloba o planejamento colaborativo e a participação cidadã.

Seu ponto de partida deve-se a uma gestão inteligente que se beneficia do uso de tecnologias como big data, análises preditivas, dispositivos de IoT, conseguindo assim, fornecer um valor agregado para a gestão que as capacita na resolução dos desafios urbanos, com políticas públicas mais assertivas, baseadas em dados e com capacidade analítica para responder mais prontamente a situações de emergência ou calamidade, escalonando serviços conforme a demanda e principalmente promovendo a participação ativa do cidadão na gestão pública.

ALGUNS EXEMPLOS DE SERVIÇOS QUE PODEM BENEFICIAR A GESTÃO E O CIDADÃO.

- Pagamento por transporte público de forma digital;
- Sinais de trânsito inteligentes;
- Manutenção previsível da infraestrutura de transporte;
- Informações em tempo real sobre transporte público;
- Detecção e controle de vazamentos de água;
- Acompanhamento do consumo de água;
- Monitoramento da qualidade da água;
- Otimização da rota de coleta de resíduos;
- Luzes de rua inteligentes;
- Sistemas de alerta precoce de desastres;
- Inspeções de construção baseadas em dados;
- Gestão de multidões;
- Mapeamento de crimes em tempo real;
- Informações em tempo real sobre a qualidade do ar;
- Monitoramento de doenças infecciosas;
- Sistemas integrados de gerenciamento de fluxo de pacientes.

E COMO INTEGRAR ESSES SERVIÇOS?

O principal é ter uma **infraestrutura de conectividade** (fibra óptica, satélite, rádio, 3G, 4G, 5G, redes sem fio) em toda a cidade e para todo cidadão, sem essa conectividade nada seria viável, pois os dados não poderiam fluir pela cidade.

Dispositivos de IoT, são eles os principais mecanismos de coleta e distribuição de dados, muitas vezes em tempo real, servindo para monitorar os vários serviços públicos, como, por exemplo, gestão de trânsito, segurança, qualidade do ar, gestão de resíduos, sistemas de iluminação etc.

Um **centro integrado** que possa administrar, armazenar e analisar todos os dados enviados pelos dispositivos IoT ou de qualquer outra fonte, permitindo uma visão unificada de todas as áreas da cidade, acelerando assim a tomada de decisões.

Interfaces de comunicação do governo com o cidadão, seja através de aplicativos móveis, plataformas web ou outros serviços on-line, sempre tendo em mente a inclusão digital.

PARTICIPAÇÃO CIDADÃ E INCLUSÃO DIGITAL.



Conforme Milani (2008), a introdução da participação dos usuários na produção das políticas permite que se promova a criação de uma rede de política pública, que realize a articulação entre setores públicos, privados e organizações da sociedade civil, que tenham interesses comuns e vantajosos para a sociedade.

Não é suficiente apenas ter a vontade de envolver os cidadãos na elaboração de políticas, é essencial também motivá-los a participar ativamente.

A adoção de serviços on-line deve ser inclusiva, garantindo que todos os meios de participação popular sejam acessíveis, levando em consideração as necessidades de pessoas com deficiências, idosos e grupos marginalizados.

A administração pública deve além de motivar o interesse dos cidadãos, criar um ambiente onde possa ocorrer essa interação, seja ele físico ou virtual. A participação cidadã requer a adoção de estratégias e ferramentas para envolver efetivamente a população nas decisões municipais.

SE TRAZ TANTO BENEFÍCIOS POR QUE TODAS AS CIDADES NÃO ADOTAM?

Por refletir mudanças substanciais no formato dos procedimentos e processos existentes na administração, bem como na interação com o cidadão, resultando em uma mudança de paradigma, alguns desafios deverão ser enfrentados e superados pela administração pública antes de começar a caminhada em direção a uma gestão baseada em evidências.

Embora existam desafios, muitas cidades estão, de fato, adotando gradualmente soluções inovadoras, usando tecnologia, para abordar problemas específicos. A conscientização sobre os benefícios potenciais e a superação desses desafios pode aumentar a adoção e transformação das cidades em ambientes mais inteligentes.

APOIO POLÍTICO.

Como a maioria das cidades brasileiras possuem orçamento limitado, dando prioridade em resolver problemas mais urgentes ou que tenha resultado mais imediato, permitindo assim, maior visibilidade da gestão perante a opinião pública, torna-se fundamental o apoio das lideranças locais, pois são elas que definem a visão estratégica da cidade.

Com o apoio das lideranças, a alocação e priorizando de recursos orçamentários e humanos podem ser facilitados; colaboração interna entre as secretarias no compartilhamento de dados; facilitar a compatibilização desses projetos no aspecto legal.

Um aspecto relevante desse apoio seria a institucionalização de projetos dessa natureza, onde poderiam ter seu risco de serem descontinuados com a mudança de liderança diminuídos.

QUADRO TÉCNICO LIMITADO.

Segundo Bouskela et al., (2016), um elemento fundamental diz respeito às considerações relacionadas aos recursos humanos, mais precisamente à constituição de uma equipe qualificada, de natureza multidisciplinar e constantemente atualizada, apta a viabilizar um projeto de big data.

Projetos de big data exigem conhecimentos em diferentes áreas e especialidades, tornando necessário uma equipe multidisciplinar o que dificulta muito a criação da mesma, principalmente em cidades de pequeno porte.

Caso existam dificuldades na criação de uma equipe interna da prefeitura, a criação de parcerias governo-academia ou governo-empresas pode ser uma solução.

A academia possui um notável potencial no desenvolvimento de soluções inovadoras, já as empresas frequentemente têm acesso a recursos financeiros, infraestrutura tecnológica e em alguns casos projetos de big data já bem definidos.

FALTA DE INTEGRAÇÃO E PADRONIZAÇÃO DOS DADOS DISPONÍVEIS.

Em muitos casos ocorre uma resistência à mudança ou falta de compreensão sobre como os dados podem realmente otimizar processos e decisões. Como as práticas antigas da administração pública sofre com a falta de integração e padronização quando o assunto são os dados, tornando difícil uma visão precisa de onde extrair as informações necessárias.

A política de dados abertos é uma boa forma para solucionar a falta de integração e padronização dos dados, pois o governo pode ter todos os seus dados publicados em um único local, facilitando a adoção de uma governança de dados e simplificando o seu uso pela gestão ou por terceiros.

JURÍDICOS E REGULATÓRIOS.

Além das leis de privacidade de dados como a Lei Geral de Proteção de Dados (LGPD), os projetos de big data podem estar sujeitos a leis e regulamentações específicas de setores, como saúde, educação e finanças. O cumprimento dessa legislação pode ser algo complexo, podendo, de acordo com a jurisdição exigir um esforço adicional para garantir que os projetos estejam em conformidade legal.

O grande desafio dessas legislações é equilibrar a proteção da privacidade e o uso de dados pessoais por terceiros, alguns princípios devem ser estabelecidos para garantir a transparência e o uso legal desses dados.

O consentimento da pessoa de forma clara para o uso de seus dados pessoais, esclarecendo a finalidade específica onde o dado será usado é uma forma de promover o uso responsável dos dados pessoais por terceiros.

TECNOLÓGICO E INFRAESTRUTURA.

A infraestrutura e a tecnologia desempenham um papel fundamental no sucesso de iniciativas de big data, muitos projetos podem fracassar quando a administração não possui uma visão clara das necessidades e do que realmente existe de concreto na cidade em termos tecnológicos.

O levantamento e avaliação da capacidade dos recursos computacionais e de infraestrutura de conectividade existentes para estabelecer se são adequados para sustentar as necessidades do projeto é de extrema importância.

Com base no levantamento possíveis gargalos ou limitações podem ser identificados. Faça um plano para atualizar ou expandir os recursos de TIC conforme necessário, estimando os custos associados às atualizações ou expansões necessárias.

Vale lembrar que a disponibilização cada vez maior de serviços online impulsiona um aumento na tentativa de invasões nesses serviços. Segurança também é crucial.

MINHA CIDADE É PEQUENA, INFRAESTRUTURA E TECNOLOGIA ESTÃO FORA DO MEU ORÇAMENTO.

Ok, mas vamos ponderar alguns aspectos relevantes, conectividade é um aspecto imprescindível em qualquer aspecto ligado a tecnologia na atualidade, esse possa ser o obstáculo mais difícil de ultrapassar.

Mas lembre comece pequeno pensando grande. Muitos projetos podem ser implantados ou até mesmo serem usados como aprendizado para fornecer experiência para a sua equipe.

Claro que uma equipe técnica será necessária, lembrando que parcerias podem ser realizadas com universidades ou startups.

VAMOS FALAR UM POUCO DE TECNOLOGIA.

As duas principais linguagem de programação usadas atualmente em análises são a linguagem R e Python, ambas são open source, basta baixar e começar a brincar.

Ambas as linguagem também contam com um ambiente de desenvolvimento totalmente on-line.

Mas o que seria isso? Não tenho computador com configurações suficientes ou estou com dificuldades na instalação e configuração dessas linguagens. Sem problemas, existem essas soluções:

Posit Cloud, permite que você acesse o poderoso conjunto de ferramentas de ciência de dados em R diretamente em seu navegador – sem necessidade de instalação ou configuração complexa.

<https://posit.cloud/>



Google Colab, é um serviço da linguagem Python que não requer configuração para ser usado e fornece acesso gratuito a recursos de computação, especialmente adequado para aprendizado de máquina, ciência de dados e educação. <https://colab.google/>



FALANDO UM POUCO MAIS DE TECNOLOGIA.

Sabemos que em relação aos dados existem várias lugares onde podemos baixá-los, quando falamos de política de dados abertos. Agora caso você não tenha os dados na sua própria base, existe uma iniciativa muito interessante chamada Base de Dados, onde são disponibilizados vários dados de bases governamentais com o tratamento já realizado desses dados, facilitando as análises.

<https://basedosdados.org/>



O Banco Interamericano de Desenvolvimento disponibiliza vários projetos já implementados em governos, todos com o código aberto, podendo ser baixados e modificados para atender as necessidades específicas de cada cidade. <https://code.iadb.org/pt>



Caso o armazenamento seja um problema existe o BigQuery, da Google, é um banco de dados sem servidor que funciona em nuvens e pode ser escalonado de acordo com seus dados, além de usar aprendizado de máquina de forma integrada. Atualmente fornece 10 GB de armazenamento e 1 TB de consultas gratuitas por mês.

<https://cloud.google.com/bigquery>



INDO UM POUCO ALÉM, ATÉ AS NUVENS.

Sabemos que em cidades de pequeno porte, a quantidade de usuários dos serviços online da prefeitura podem não justificar a aquisição e manutenção de uma infraestrutura de TIC própria.

Em vez de adquirir e manter infraestrutura de TIC localmente, os órgãos públicos podem usar esses serviços em nuvem para atender às suas necessidades em projetos de big data.

A computação em nuvem oferece várias vantagens, incluindo escalabilidade, segurança, flexibilidade, economia de custos e acessibilidade.

EXISTEM VÁRIOS TIPOS DE NUVEM.

nuvem pública: infraestrutura de nuvem onde os recursos de computação, armazenamento e rede são compartilhados entre várias organizações ou indivíduos.

nuvem privada: infraestrutura de nuvem dedicada a uma única organização ou empresa, oferecendo maior controle, segurança e personalização dos recursos, permitindo que a organização atenda a requisitos específicos de conformidade e segurança;

nuvem híbrida: combinação de nuvem pública e nuvem privada que permite que os dados e aplicativos sejam compartilhados entre elas,

as organizações podem mover cargas de trabalho entre nuvens públicas e privadas de acordo com as necessidades, proporcionando flexibilidade e escalabilidade;

nuvem comunitária: infraestrutura de nuvem compartilhada por várias organizações com interesses comuns ou requisitos de conformidade, permitindo que organizações com necessidades semelhantes colaborem e compartilhem recursos em uma nuvem dedicada a uma comunidade específica, oferece controle e segurança compartilhados, ao mesmo tempo em que atende aos requisitos exclusivos da comunidade, é uma alternativa interessante que pode ser usada por organizações governamentais.

EXISTEM VÁRIAS FORMAS DE ENTREGA DE RECURSOS E SERVIÇOS.

infraestrutura como serviço (IaaS): é o modelo que mais se aproxima de uma infraestrutura local, seria como ter sua própria infraestrutura virtualizada, podendo incluir servidores virtuais, armazenamento, redes e outros componentes de infraestrutura;

plataforma como serviço (PaaS): esse modelo oferece um ambiente completo para o desenvolvimento de aplicações na nuvem, os desenvolvedores conseguem criar, gerenciar e hospedar suas aplicações sem se preocupar com atualizações de softwares usados para o desenvolvimento e nem a infraestrutura necessária para isso;

software como serviço (SaaS): neste modelo, um software é entregue pela internet como um serviço, os clientes o acessam através da internet, sem a, seria como alugar um software pela internet.

EXISTEM VÁRIOS PROVEDORES DE SERVIÇOS EM NUVEM.

AMAZON WEB SERVICES (AWS)

<https://aws.amazon.com/pt>



MICROSOFT AZURE

<https://azure.microsoft.com/pt-br>



GOOGLE CLOUD

<https://cloud.google.com/>



A escolha de um provedor de nuvem deve ser feita cuidadosamente, existem várias no mercado, mas deve ser sempre levado em consideração os requisitos específicos de cada projeto, os custos e as considerações de segurança e conformidade.

Lembrando que uma solução em nuvem não irá ser viável se a cidade não dispor de uma infraestrutura de conectividade que satisfaça as demandas de acesso com o provedor em nuvem, além de possui uma equipe técnica qualificada que consiga implementar e gerenciar os recursos oferecidos pelos provedores.

E O CUSTO DESSES SERVIÇOS EM NUVEM.

No caso de uma migração para computação em nuvem os gastos ficariam variados de acordo com os recursos utilizados em relação a processamento, armazenamento, tempo de uso, dados utilizados e em alguns casos o suporte técnico.

Também pode ocorrer uma mudança na estrutura da equipe de TIC, devido ao provedor do serviço em nuvem assumir a responsabilidade por alguns recursos, principalmente os relacionados a administração da infraestrutura.

Mas vale lembrar que vários provedores oferecem serviços de forma totalmente gratuitos ou com períodos de gratuidade de 12 meses, além de oferecer ferramentas de controle de gastos, por exemplo, a AWS possui a AWS Pricing Calculator que é uma calculadora de preços online que permite que os clientes estimem os custos dos serviços da AWS com base em suas necessidades específicas e o AWS Budgets, onde os clientes podem criar orçamentos personalizados com base em suas necessidades e metas de gastos, especificando o valor máximo que desejam gastar em um período específico, como um mês.

FIQUE ATENTO!

Um grande problema encontrado para a utilização em computação em nuvem é a ausência de regulamentação. Esses entraves normativos criam insegurança em relação a viabilidade de contratação de provedores devido as incertezas nas questões relacionadas a modalidade de adoção, precauções necessárias, riscos associados, entre outros.

NADA MELHOR QUE UM EXEMPLO PRÁTICO



Iremos construir dois mapas com informações sobre furtos de veículos. Os dados foram coletados do site do IBGE e da Secretaria de Segurança Pública do Estado de São Paulo, referente aos furtos de veículos nos meses de dezembro de 2022 e janeiro de 2023, nos sites:

<https://www.ibge.gov.br/geociencias/organizacao-do-territorio/malhas-territoriais/15774-malhas.html>



Acesse esse link
usando o QR code

<https://www.ssp.sp.gov.br/transparenciassp/>



Acesse esse link
usando o QR code

Escolhemos o município de Morro Agudo do estado de São Paulo, que faz parte da Microrregião de São Joaquim da Barra, com população de 32.968 habitantes em 2019.

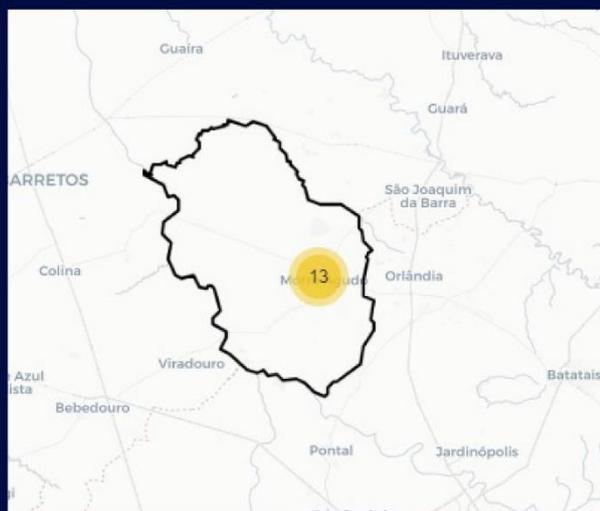
A escolha do município e do tipo de ocorrências foi aleatória, mas levando em conta um município com população menor que 50 mil habitantes. Poderíamos ter usado qualquer município de São Paulo, bem como qualquer ocorrência disponibilizada pela Secretaria de Segurança Pública.

Vale lembrar que existem muitas informações governamentais disponíveis sobre outros temas como saúde, educação, economia, etc.

O código foi escrito em **Python** e foi utilizado a plataforma do **Google Colab**, serão gerados mapas interativos que podem ser exportados no formato html e poderão servir de evidências para o gestor público tomar providências para inibir ou diminuir os casos de furtos de veículos.



Veremos que na Microrregião de São Joaquim da Barra ocorreram 40 ocorrências durante o período analisado.



Já no município de Morro Agudo, no mesmo período, ocorreram 13 furtos de veículos o que corresponde a 32,5% do total das ocorrências da Microrregião.



E quando analisados com mais detalhes, pode ser verificado que 6 das ocorrências são bem próximas.

Algumas medidas simples poderiam ser tomadas por parte da prefeitura, como a instalação de câmeras de monitoramento nessa área para inibir essas ações.

A instalação de câmeras de monitoramento em áreas públicas é um exemplo de uma política pública voltada para a segurança urbana, buscando melhorar a segurança pública na comunidade, utilizando uma abordagem preventiva.

Além da instalação de câmeras, a prefeitura também pode considerar outras medidas complementares, como melhorar a iluminação pública e a promoção de programas comunitários que incentivem a participação dos moradores na segurança local.

A escolha de medidas específicas dependerá das necessidades e características da comunidade em questão. Em última análise, o objetivo é melhorar a qualidade de vida dos cidadãos e promover um ambiente seguro.

Exemplos Prático

✓ Será realizada uma análise real com os dados de furto de veículos do município de Morro Agudo.

Morro Agudo é um município brasileiro do estado de São Paulo, que faz parte da Microrregião de São José do Rio Preto, composta por 9 municípios. Sua população em 2019 era de 32.568 habitantes.

Serão utilizados dados oficiais de IBGE e da Secretaria da Segurança Pública de São Paulo. Os dados serão dos meses de dezembro de 2022 e janeiro de 2023, referentes a furto de veículos.

Instalando as bibliotecas:

```
[1]: !pip install pandas
!pip install matplotlib
!pip install geopandas
!pip install folium
!pip install folium.plugins
```

Como importar as bibliotecas:

```
[1]: import pandas as pd
import matplotlib.pyplot as plt
import geopandas as gpd
import folium
from folium.plugins import HeatMap, LineString, Polygon
from folium.plugins import MarkerCluster
```

Acesse o código da aplicação prática através do QR code:



Considerando o momento, o grande volume de dados gerados, principalmente pela população; o aumento da possibilidade de coleta de dados em tempo real através de dispositivos IoT; as novas técnicas de análises, empregando técnicas de aprendizado de máquinas, que utilizam esses dados e são capazes de processá-los em curto tempo na tomada de decisões mais assertivas; o barateamento do custo em serviços em nuvem, que alguns provedores ofertam seus serviços de forma gratuita ou com carência de tempo para uso; linguagens open source como R e Python que são as mais utilizadas nessas análises; as parcerias que podem ser criadas. Ao integrar esses elementos ou parte deles, as cidades moldam o seu futuro de maneira mais sustentável e inteligente.

“

**A MELHOR MANEIRA DE
SE PREVER O FUTURO
É CRIANDO-O.**

”

PETER F. DRUCKER

REFERÊNCIAS BIBLIOGRÁFICAS

BETSER, Joseph; BELANGER, David. Architecting the enterprise via big data analytics. In: LIEBOWITZ, Jay. Big data and business analytics. Boca Raton: Taylor & Francis Group, LLC, 2013.

BOUSKELA, Mauricio, et al. Caminho para as Smart Cities. Banco Interamericano de Desenvolvimento (2016).

CERDEIRA, Pablo de Camargo; MENDONÇA, Marcus Mentzingen de; LAGOWSKA, Urszula Gabriela. Políticas públicas orientadas por dados: os caminhos possíveis para governos locais. BID, 2020.

DATA SCIENCE ACADEMY. Deep Learning Book Disponível em: <https://www.deeplearningbook.com.br/o-que-sao-redes-neurais-artificiais-profundas/>. Acesso em: 07 set. 2023.

DE LUCA, Cristina; BASSI, Silvia. Potencializando o uso de big data para cidades inteligentes: um guia estratégico para gestores. 2023.

DINIZ. O governo eletrônico no Brasil: perspectiva histórica a partir de um modelo estruturado de análise. Revista de Administração Pública-RAP, Rio de Janeiro, n. 43, v. 1, p. 23-48, jan./fev. 2009. Disponível em: http://www.scielo.br/scielo.php?pid=S0034-76122009000100003&script=sci_abstract&tling=pt. Acesso em: 05 jun. 2023.

KIM, Gang-Hoon, Silvana TRIMI, and Ji-Hyong CHUNG. Big-data applications in the government sector. Disponível em: <<https://linkedkey.com/wp-content/uploads/2016/05/Big-Data-Applications-in-the-Government-Sector.pdf>>. Acesso em: 18 jul. 2023.

MACHADO, F. N. R. Big Data: o Futuro dos Dados e Aplicações, 1. ed. São Paulo: Saraiva, 2018

MILANI, C. R. S. O princípio da participação social na gestão de políticas públicas locais: uma análise de experiências latino-americanas e europeias. Revista de Administração Pública, v. 42, n. 3, p. 551–579, jun. 2008.

OPEN KNOWLEDGE INTERNATIONAL. Disponível em: <<https://ok.org.br/dados-abertos>>. Acesso em: 07 jul. 2023.

SCHÖNBERGER-MAYER, Viktor; CUKIER Kenneth. Tradução Paulo Palzonoff Junior. Big data: como extrair volume, variedade, velocidade e valor da avalanche de informação cotidiana. Rio de Janeiro: Elsevier, 2013.

TAURION, Cezar. Big data. Rio de Janeiro: Brasport, 2013.

TRIBUNAL DE CONTAS DA UNIÃO – TCU. Secretaria de Fiscalização de Tecnologia da Informação. 5 motivos para a abertura de dados na Administração Pública. Brasília, 2015. Disponível em: <<http://portal3.tcu.gov.br/portal/pls/portal/docs/2689107.PDF>>. Acesso em: 11 jul. 2023.