

## **Análise de Dados na Investigação do Impacto do Índice de Desenvolvimento Humano Municipal em Resultados Eleitorais**

Nilton Cesar Sacco, Eduardo J. Huerta Yero, M. Carmo Nicoletti  
UNIFACCAMP

Campo Limpo Paulista - SP, Brasil

[ncesar\\_and@hotmail.com](mailto:ncesar_and@hotmail.com), [edu.huerta@gmail.com](mailto:edu.huerta@gmail.com), [carmo@cc.faccamp.br](mailto:carmo@cc.faccamp.br)

**Resumo.** Este artigo descreve parte de um trabalho de pesquisa sendo conduzido na área de análise de dados, que tem como objetivo investigar possíveis relações entre determinados indicadores que caracterizam aspectos de eleitores e municípios brasileiros, com o resultado obtido em uma eleição em nível federal. Um dos indicadores já escolhido foi o Índice de Desenvolvimento Humano Municipal (IDHM), que busca caracterizar numericamente o nível de desenvolvimento e a qualidade de vida oferecida por municípios brasileiros. O IDHM é derivado do Índice de Desenvolvimento Humano (IDH), um indicador global considerado uma das métricas mais amplamente aceitas para representar o status de desenvolvimento de um país. O trabalho sendo realizado pretende analisar os resultados das eleições federais à presidência, ocorridas em 2018, por meio de exames dos índices IDHM, bem como do número de habitantes de cada município.

**Palavras-chave:** extração de dados, análise de dados, dados eleitorais, IDH, IDHM.

**Abstract.** This article describes part of a research work being conducted in the area of data analysis, which aims to investigate possible relationships between certain indicators that characterize aspects of Brazilian voters and municipalities, with the result obtained in an election at the federal level. One of the indicators already chosen was the Municipal Human Development Index (MHDI), which seeks to numerically characterize the level of development and quality of life offered by Brazilian municipalities. The MHDI is derived from the Human Development Index (HDI), a global indicator and one of the most widely accepted metrics to represent a country's development status. The work being carried out intends to analyze the results of the federal presidential elections, which took place in 2018, by examining the MHDI indices, as well as the number of inhabitants of each municipality.

**Keywords:** data extraction, data analysis, electoral data, HDI, MHDI.

### **1. Introdução**

O conceito de dado praticamente permeia todo o conhecimento humano. De uma maneira simplista, dado pode ser caracterizado como observações ou, então, resultados de medições coletadas via instrumentação ou ainda, resultados de simples contagens/identificações. Outra caracterização comum e razoavelmente imprecisa é a de dado ser parte de uma informação, mas não ser informação propriamente dita.

Quando dados são organizados, interpretados ou estruturados de maneira a terem um significado ou então, alguma utilidade, são considerados informações. Já o conceito de conhecimento é bem mais abrangente e, apesar de ser intrinsecamente dependente dos conceitos de dado e informação, vai bem além, envolvendo consciência e cognição, como definido em (Wikipedia 2020), ao considerar conhecimento como uma familiaridade, consciência ou entendimento de alguém ou algo, como fatos, informações, descrições ou habilidades, adquiridos por meio da experiência, ou da educação pela percepção, descoberta ou aprendizado.

Invariavelmente os três conceitos, dado, informação e conhecimento são abordados de maneira inter-relacionada, devido à interdependência entre eles. Uma abordagem superficial que é comumente aceita e é apresentada em (Tuomi 1999), é a de dados serem tratados como fatos simples que se tornam informações, à medida que são combinados em estruturas significativas e que, subsequentemente, se tornam conhecimento à medida que as informações significativas são colocadas em um contexto e quando podem ser usadas para fazer previsões.

Na literatura existem inúmeros trabalhos abordando os três conceitos, nos mais variados contextos do conhecimento humano (ver (Tuomi 1999) (Tan et al. 2006) (Zins 2007) (Diffen 2020)). Em vários trabalhos com foco em uso/análise de dados, dados podem também ser referenciados como dados brutos i.e., dados obtidos que não passaram por um processo de tratamento para eliminar ou, possivelmente, recuperar dados com problemas (por exemplo, incompletos ou espúrios).

É fato que um volume imenso de dados brutos vem sendo produzidos diariamente, por fontes dos mais variados tipos, tais como redes sociais, órgãos governamentais, empresas estatais, empresas privadas, empresas sem fins lucrativos, bancos, universidades, meios de comunicação, sites informativos e de divulgação, e muitos outros. Muitos desses dados acabam se tornando obsoletos ao longo do tempo devido, principalmente, à sua natureza volátil.

Muitas organizações, que ainda não estão sensibilizadas com relação à importância que os dados que produzem têm no planejamento de suas atividades, acabam ignorando muitos dos dados produzidos. Eventualmente tais dados se tornam desatualizados ao longo do tempo e, então, são descartados. Já organizações mais ágeis e sensíveis à importância que dados desempenham em quase todas as atividades que implementam/planejamos, buscam desenvolver procedimentos que utilizam muitos dos dados produzidos para subsidiar a definição/refinamento de muitos de seus processos e, particularmente, os processos administrativos relacionados a tomadas de decisão.

Dados produzidos ou coletados servem, por exemplo, para que organizações, por meio da análise de tais dados, conheçam a opinião de seus clientes sobre seus produtos/serviços ou, então, para poderem corrigir problemas relacionados aos seus produtos/serviços. Outro exemplo da coleta e uso de tais dados é aquele realizado em hospitais quando do monitoramento de pacientes, por meio de sensores colocados em pacientes que estão sob cuidados médicos contínuos, com o objetivo de coletar informações sobre suas principais funções vitais, para acompanhamento on-line de suas situações clínicas (Wesley et al. 2003) (An et al. 2012). Também, imagens digitalizadas podem contribuir tanto para a equipe médica substanciar com mais precisão diagnósticos, quanto para acompanhar a resposta/evolução do estado do paciente aos tratamentos. Não pode ser esquecido, entretanto, que existem ainda muitas áreas em que o uso de dados passíveis de serem coletados ainda não foram devidamente exploradas, devido a dificuldades práticas envolvidas com a própria área. É o caso, por exemplo, de dados relativos a áreas relacionadas atividades ilegais, que permeiam muitas das atividades legais em comunidades sociais (Kejriwal & Szekely 2007).

Este artigo está organizado em mais 3 seções. A Seção 2 descreve brevemente os dados governamentais disponibilizados publicamente que serão usados no contexto do trabalho de pesquisa sendo realizado e comenta sobre o uso de dados disponibilizados pelo Censo Demográfico de 2010. A Seção 3 aborda inicialmente o IDH, utilizado pelo PNUD para, então, detalhar o IDHM, um dos descritores já escolhido para a caracterização de eleitores e municípios. A Seção 4 finaliza o artigo, lembrando os

objetivos da pesquisa sendo desenvolvida e lista algumas das atividades planejadas para a continuação e finalização da pesquisa.

## 2. Dados Eleitorais e Dados de Municipalidades

Hoje o Brasil se enquadra como uma fonte produtiva de dados, continuamente produzindo e disponibilizando um grande volume de dados em repositórios públicos, estruturados, que podem ser acessados pelo público em geral e utilizados para os mais variados objetivos. Algumas leis e princípios estabelecidos pelo governo brasileiro garantem a veracidade, disponibilização e o acesso de um vasto volume de dados governamentais, por qualquer cidadão. É fato, entretanto, que apenas alguns poucos sistemas computacionais, propostos e implementados por órgãos de algumas das esferas e poderes da administração pública e alguns da iniciativa privada, têm sido efetivamente utilizados para a exploração e extração de informações/conhecimento a partir de tais dados.

A disponibilização de dados governamentais ao acesso da população brasileira foi motivada, particularmente, pela iniciativa de órgãos públicos governamentais brasileiros em promover a transparência de vários processos e movimentações do dinheiro público. O conceito de transparência utilizado pelos órgãos públicos é abordado com detalhes no site do governo federal conhecido como *Portal da Transparência* (<http://www.portaltransparencia.gov.br/>), administrado pela Controladoria-Geral da União (CGU). O *Portal da Transparência* disponibiliza ao acesso público informações sobre a gestão pública e sobre a aplicação do dinheiro público no Brasil, o que possibilita que cidadãos possam monitorar o uso dos recursos federais arrecadados, particularmente aqueles advindos de impostos.

Também, foi objetivo governamental, ao disponibilizar publicamente dados referentes ao capital humano, tais como os índices de desenvolvimento, o de promover a clareza de processos eleitorais, com o intuito de coibir fraudes. Particularmente, a disponibilização de tais dados abriu uma área de investigação empírica com grande potencial para a obtenção de resultados concretos, representados por informações ou conhecimento empírico, obtidos por procedimentos estatísticos e/ou algoritmos de Aprendizado de Máquina (AM) (Bishop 2006) e Mineração de Dados (MD) (Tan et al. 2006).

A última eleição presidencial no Brasil aconteceu em 2018 e foi realizada em dois turnos; o primeiro turno em 7 de outubro e o segundo em 28 de outubro. No segundo turno das eleições de 2018 concorreram à presidência dois candidatos, um do Partido dos Trabalhadores (PT) (<https://pt.org.br/>) e o outro do Partido Social Liberal (<https://psl.org.br/>). Neste artigo cada uma das siglas, PT e PSL, representa ambos, o partido político e o candidato do partido político que concorreu à presidência.

As investigações a serem conduzidas pela pesquisa em andamento, parcialmente descritas neste artigo, investem na identificação de um conjunto de descritores de dados que, potencialmente, teriam relevância para promover uma análise comparativa dos dados referentes aos resultados do segundo turno das eleições presidenciais de 2018 e, também, de descritores que poderiam estar correlacionados com os resultados da eleição. A pesquisa contempla também o uso de dados relativos ao Censo Demográfico de 2010, realizado pelo IBGE (Instituto Brasileiro de Geografia e Estatística), que estão disponibilizados no site <https://censo2010.ibge.gov.br/>. Como informado no site, (1) no Censo 2010, mais de 190 mil recenseadores visitaram 67,6 milhões de domicílios nos 5.565 municípios brasileiros (2) informações sobre todas as etapas do Censo 2010 e,

particularmente, os resultados do Censo, estão disponibilizados publicamente. Espera-se que com o auxílio de análises estatísticas dos dados, condensados por meio do uso de descritores, padrões que relacionam os índices de desenvolvimento humano dos municípios brasileiros, com a escolha de voto do eleitorado, em cada um dos municípios considerados pelo Censo, permitam a determinação de um perfil eleitoral particular a cada um dos dois lados da disputa presidencial que aconteceu no segundo turno das eleições de 2018.

### 3. Índice de Desenvolvimento Humano Municipal

O Programa das Nações Unidas para o Desenvolvimento (PNUD) (<https://www.br.undp.org/>) é uma rede global de desenvolvimento da ONU Organização das Nações Unidas (<https://nacoesunidas.org/>). Presentemente o PNUD está estabelecido em 170 países e territórios, com a missão de alinhar seu trabalho às necessidades do país, colaborando no desenvolvimento de políticas, habilidades de liderança, capacidades institucionais, resiliência e, especialmente, erradicação da pobreza e redução de desigualdades e exclusão social.

O PNUD busca promover o desenvolvimento humano, e contemplar os objetivos de desenvolvimento sustentável nos países em que atua, por meio de investimentos em três áreas principais: (1) desenvolvimento sustentável; (2) apoio à política democrática e construção da paz e (3) resiliência ao clima e desastres. O PNUD também busca ajudar países a atrair ajuda de maneira efetiva e, em suas atividades, promove igualdade de gênero e a proteção dos direitos humanos. A estratégia de atuação do PNUD no Brasil contempla, entre várias outras metas, a de investir em áreas vulneráveis e populações de baixo IDH. O conceito de desenvolvimento humano, como adotado pelo PNUD, é definido como um processo de ampliação das escolhas das pessoas para que tenham capacidades e oportunidades para se tornarem o que desejam ser. A composição do IDH leva em consideração três áreas:

(1) *saúde*, medida pela expectativa de vida;

(2) *educação*, medida por:

(2a) média de anos de educação de adultos, que é o número médio de anos de educação recebidos, durante a vida, por pessoas a partir de 25 anos e

(2b) a expectativa de anos de escolaridade para crianças na idade de iniciar a vida escolar, que é o número total de anos de escolaridade que uma criança na idade de iniciar a vida escolar pode esperar receber, se os padrões prevalecentes de taxas de matrículas específicas por idade permanecerem os mesmos durante a vida da criança.

(3) *padrão de vida*: medido pela Renda Nacional Bruta (RNB) per capita, expressa em poder de paridade de compra (PPP), em dólar, tendo 2005 como ano de referência.

Uma especialização do IDH, conhecida como IDHM é uma medida que também agrega indicadores das três perspectivas do desenvolvimento humano: longevidade, educação e renda. O valor que esse índice pode assumir está confinado ao intervalo [0 1]. Quanto mais próximo de 1 estiver o seu valor, maior o desenvolvimento humano municipal. Em seu maior valor o índice representa a oportunidade de viver uma vida longa e saudável, de ter acesso ao conhecimento e ter um padrão de vida que garanta as necessidades básicas.

Além de contemplar as três perspectivas do IDH, IDHM brasileiro adequa a metodologia global utilizada para o estabelecimento do IDH, ao contexto brasileiro e

uso dos indicadores nacionais disponíveis. Embora meçam os mesmos fenômenos, os indicadores levados em conta no IDHM são mais adequados para avaliar o desenvolvimento dos municípios brasileiros. O IDHM é composto por três componentes: IDHM Longevidade, IDHM Educação e IDHM Renda, como descrito a seguir.

(1) *IDHM Longevidade*: representa numericamente a expectativa de vida (em anos), ao nascer, calculada por método indireto a partir de dados de censos Demográficos do IBGE. Representa o número médio de anos que as pessoas viveriam a partir do nascimento, mantidos os mesmos padrões de mortalidade observados no ano de referência.

(2) *IDHM Educação*:

O nível educacional da população adulta (PA) que chamaremos de APEL, é medido como a porcentagem de pessoas como idade maior ou igual a 18 anos que concluíram o ensino fundamental. O nível educacional da população jovem (PJ) é medido como a média aritmética de quatro indicadores diferentes: O valor de IDHM Educação é então calculado como a média geométrica de PA e PJ, com PA ponderado por 1 e PJ ponderado por 2.

(3) *IDHM Renda*: medido como a renda municipal per capita i.e., a renda média de cada residente de determinado município. É a soma da renda de todos os residentes dividida pelo número de pessoas que moram no município, incluindo crianças e pessoas sem registro de renda. Os dados são obtidos do Censo Demográfico do IBGE.

Diante da atual situação, em que o Brasil passa por um longo período de turbilhão político, econômico e social, o projeto de pesquisa em andamento está explorando as fontes de dados, como uma tentativa de entender se a variação desses indicadores, que segundo estudos alteram a qualidade de vida da população, podem influenciar a opinião do eleitor permitindo visualizar seu reflexo nos municípios brasileiros em âmbito nacional.

#### **4. Continuação do Trabalho e Considerações Finais**

Como brevemente comentando anteriormente, os estudos, o levantamento bibliográfico contínuo e as investigações até agora conduzidos têm como objetivo investigar, por meio de análises estatísticas a serem conduzidas em resultados do segundo turno das eleições presidências de 2018, padrões que relacionam valores do IDHM relativos aos municípios brasileiros, com a escolha de voto do eleitorado, em cada um desses municípios. Possivelmente tal investigação permitirá o estabelecimento de perfis de eleitores que votaram em PT ou em PSL. Tendo em vista o objetivo da pesquisa, as seguintes atividades estão previstas para a continuidade e finalização do trabalho.

(1) Leitura e avaliação das considerações e críticas com relação à real utilidade dos índices de desenvolvimento humano IDH e IDHM que foram publicadas em (McGillivray 1991) e (Sagar & Najam 1998).

(2) coleta de dados de interesse à pesquisa: determinação da sua origem, veracidade e confiabilidade. Dentre os dados cogitados estão os referentes às eleições presidenciais de 2018 e os referentes aos índices de desenvolvimento humano por município.

(3) Implementação de todo o processo de preparação dos dados incluindo as etapas utilizadas desde a importação até a geração dos gráficos e as conclusões preliminares.

Dados a serem utilizados deverão também passar por um pré-processamento com o objetivo de eliminar aqueles com ruído e incompletos.

(4) Familiarização com o software Exploratory (Exploratory 2020), previsto para ser utilizada na fase inicial do pré-processamento de dados

(5) identificação dos principais conceitos da Estatística que serão considerados para análises numéricas a serem conduzidas.

(6) exploração no uso de algoritmos de Aprendizado de Máquina (AM) com vista à indução de conhecimento associado a perfis de usuários, com base em descritores baseados nos índices considerados.

**Agradecimentos** O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior-Brasil (CAPES)-Código de Financiamento 001.

## Referências

An, Y.; Khare, R.; Song, I.-Y.; Hu, X. (2012) Data exploration and knowledge discovery in a patient wellness tracking (PWT) system at a nurse-managed health services center, In: Proc. of The ACM International Health Informatics Symposium (IHI 2012), USA, pp. 661–665.

Bishop, C. (2006) Pattern Recognition and Machine Learning, Springer Science+Business Media.

Diffen (2020) Data vs Information, Acessado: Junho/2020, [https://www.diffen.com/difference/Data\\_vs\\_Information](https://www.diffen.com/difference/Data_vs_Information).

Exploratory, Inc. *Exploratory*. <https://exploratory.io>.

Kejriwal, M.; Szekely, P. (2007) Data mining in unusual domains with information-rich knowledge graph construction, inference and search, Talk presented at The KDD 2007.

McGillivray, M. (1991) The human development index: Yet another redundant composite development indicator? *World Development*, v. 19, no. 10, pp. 1461–1468.

Sagar, A. D.; Najam, A. (1998) The human development index: a critical review, *Ecological Economics*, v. 25, pp. 249–264.

Tan, P.-N.; Steinbach, M.; Kumar, V. (2006) *Introduction to Data Mining*, Pearson Education, Inc.

Tuomi, I. (1999) Data is more than knowledge: implications of the reversed knowledge hierarchy for knowledge management and organizational memory, *Journal of Management Information Systems*, v. 16, no. 3, pp. 103–117.

Wesley et al. (2003) Monitoring sedation status over time in ICU patients, *Journal of the American Medical Association (JAMA)*, v. 289, no. 22, pp. 2983–2991.

Wikipedia Knowledge Acessado: 08/junho/2020. <https://en.wikipedia.org/wiki/Knowledge>

Zins, C. (2007) Conceptual approaches for defining data, information, and knowledge, *Journal of the American Society for Information Science and Technology*, v. 58, no. 4, pp. 479–493.