

UNIVERSIDADE FEDERAL DE JUIZ DE FORA
FACULDADE DE ECONOMIA
GRADUAÇÃO EM CIÊNCIAS ECONÔMICAS

THOMAS NATANAEL PASSOS DA COSTA

**UM ESTUDO DOS MUNICÍPIOS DE MINAS GERAIS ATRAVÉS DA APLICAÇÃO
DOS MAPAS AUTO-ORGANIZÁVEIS DE KOHONEN SOBRE O IDHM**

Juiz de Fora
2026

THOMAS NATANAEL PASSOS DA COSTA

**UM ESTUDO DOS MUNICÍPIOS DE MINAS GERAIS ATRAVÉS DA APLICAÇÃO
DOS MAPAS AUTO-ORGANIZÁVEIS DE KOHONEN SOBRE O IDHM**

Monografia apresentada ao curso de
Ciências Econômicas da Universidade
Federal de Juiz de Fora, como requisito à
obtenção do título de bacharel.

Orientador: Prof. Alexandre Zanini

Juiz de Fora

2026

RESUMO

O estudo faz a análise do desenvolvimento socioeconômico dos municípios de Minas Gerais pela aplicação dos Mapas Auto-Organizáveis de Kohonen (MAK) nos componentes do Índice de Desenvolvimento Humano Municipal (IDHM). Buscando encontrar relações entre os dados não captadas pelo IDHM, como discrepâncias entre os valores de educação, renda e longevidade.

Os resultados revelaram nove clusters distintos, com perfis socioeconômicos claros e forte padrão de distribuição regional: os clusters de melhor desempenho concentram-se nas regiões Metropolitana de Belo Horizonte, Triângulo Mineiro e Sul de Minas, enquanto os de pior desempenho predominam no Norte e Nordeste do estado, especialmente no Vale do Jequitinhonha e Mucuri. O estudo demonstra o potencial do MAK para identificar relações entre os dados, e segmentar os municípios mineiros por suas características.

Palavras-chave: Mapa Auto-Organizável de Kohonen; IDHM; desenvolvimento humano; desigualdades regionais; Minas Gerais.

ABSTRACT

This study analyzes the socioeconomic development of the municipalities of Minas Gerais, through the application of Kohonen Self-Organizing Maps (KAM) to the components of the Municipal Human Development Index (MHDI). It seeks to find relationships between data not captured by the MHDI, such as discrepancies between the values of education, income, and longevity.

The results revealed nine distinct clusters, with clear socioeconomic profiles and a strong pattern of regional distribution: the best-performing clusters are concentrated in the Metropolitan Region of Belo Horizonte, the Triângulo Mineiro region, and Southern Minas Gerais, while the worst-performing clusters predominate in the North and Northeast of the state, especially in the Jequitinhonha and Mucuri Valleys. The study demonstrates the potential of the Kohonen Self-Organizing Map (KAM) to identify relationships between data and segment municipalities in Minas Gerais by their characteristics.

Keywords: Kohonen Self-Organizing Map; HDI; human development; regional inequalities; Minas Gerais.

AGRADECIMENTOS

Eu agradeço primeiramente a minha família, a minha mãe e meu pai, meus avós e meus tios. Que tanto trabalharam e se esforçaram para que eu pudesse ter a oportunidade de estudar em uma universidade federal, e que me apoiaram em todas as etapas desse processo.

Agradeço também aos professores, amigos e todos que passaram pelo meu caminho, nestes anos de universidade, e contribuíram para o fim desta jornada.

LISTA DE ABREVIATURAS E SIGLAS

IDH – Índice de Desenvolvimento Humano

IDHM – Índice de Desenvolvimento Humano Municipal

MAK – Mapa Auto-Organizável de Kohonen

PNUD – Programa das Nações Unidas para o Desenvolvimento

PIB – Produto Interno Bruto

IPEA – Instituto de Pesquisa Econômica Aplicada

RNA – Rede Neural Artificial

LISTA DE GRÁFICOS

Gráfico 1 - Histogramas das variáveis analisadas.....	23
Gráfico 2 - Dispersão com IDHM Educação, Renda e Longevidade.....	24

LISTA DE TABELAS

Tabela 1 - Variáveis de interesse.....	21
Tabela 2 - Medidas básicas das variáveis.....	22

SUMÁRIO

1 INTRODUÇÃO.....	9
2 REFERENCIAL TEÓRICO.....	11
3 METODOLOGIA.....	16
4 ANÁLISE DOS DADOS E RESULTADOS.....	19
5 CONCLUSÃO.....	28
REFERÊNCIAS.....	30

1 INTRODUÇÃO

Os indicadores, e as metodologias usadas para criá-los, são áreas importantes dos estudos das ciências sociais. E o são, por três motivos claros: a necessidade de medir o desempenho, e de direcionar, políticas públicas; a necessidade de legitimar com dados empíricos as reivindicações da sociedade, e as políticas governamentais; e a democratização de informações sobre a realidade social (KAYANO & CALDAS, 2001).

Um indicador importante é o Índice de Desenvolvimento Humano (IDH), indicador que sintetiza saúde, educação e renda, em um único valor. E tem como objetivo representar o nível de desenvolvimento humano e bem-estar. A saúde é representada através da esperança de vida à nascença; a educação é representada por duas métricas: a média de anos de escolaridade e os anos de escolaridade esperados; por fim, o rendimento nacional bruto per capita, é a métrica utilizada para representar a renda (PNUD, 2010).

Como uma forma de adaptação do IDH ao contexto brasileiro, e compatibilização com os indicadores nacionais, surgiu o Índice de Desenvolvimento Humano Municipal (IDHM). O IDHM abrange as mesmas dimensões sociais do IDH, porém é mais adequado para a avaliação do desenvolvimento dos municípios brasileiros (PNUD, 2013). É composto pelo IDHM Longevidade, IDHM Educação e IDHM Renda, representando as três dimensões levadas em conta pelo IDH, as principais adaptações foram feitas nos indicadores de renda e educação (GUIMARÃES & JANUZZI, 2005). Categoriza os municípios em categorias de desenvolvimento, de acordo com sua pontuação no índice: muito baixo (0-0,499), baixo (0,500-0,599), médio (0,600-0,699), alto (0,700-0,799) e muito alto (0,800-1,000) (PNUD, 2013).

Porém, existe um problema quanto ao uso do IDHM para cumprir um de seus propósitos, o direcionamento de políticas públicas. É possível que a análise do IDHM, confunda dois tipos de situações completamente diferentes: os municípios abastados que possuem condições sociais precárias, e os municípios sem uma alta renda, que apresentam boas condições sociais (VEIGA, 2003). No estado de Minas Gerais este é um problema particularmente comprometedor, pois é de uma realidade bastante complexa, onde o crescimento de certas regiões, resultou em diversos problemas

sociais como poluição, trânsito, criminalidade, dificuldades de acesso a serviços públicos, e etc (BATELLA & DINIZ, 2006). O uso de um método diferente de análise pode agregar ao planejamento das políticas públicas, apresentando novas relações entre os dados, que não são englobadas pelo IDHM, como as situações citadas anteriormente.

O objetivo geral deste trabalho é, utilizando os mesmos dados do IDHM, aplicar um método conhecido como Mapa Auto-Organizável de Kohonen, para classificar os municípios mineiros em categorias de desenvolvimento social. Como objetivos específicos tem-se o levantamento e a preparação dos dados, a aplicação do modelo e a análise dos resultados.

Em síntese, o propósito desta monografia é compreender, em que medida os componentes do IDHM permitem identificar perfis distintos de desenvolvimento municipal em Minas Gerais quando analisados por um método não supervisionado.

Este trabalho é justificado pela pequena quantidade de trabalhos, que utilizam o MAK como método de análise de IDH e IDHM na literatura brasileira. Em específico no estado de Minas Gerais, este seria, até onde foi possível identificar, o primeiro trabalho a utilizar tal metodologia, que pode auxiliar o embasamento de políticas públicas.

O trabalho é dividido em 5 seções, sendo composto pela introdução, referencial teórico, metodologia, análise dos dados e resultados e conclusão. A primeira se trata de uma apresentação inicial do IDH, IDHM e do problema abordado pela monografia, da inadequação do IDHM em captar certas relações subjacente nos dados; a segunda aprofunda e amplia o conhecimento dos indicadores, e dos problemas do IDHM, além de incluir o MAK nesta discussão; a terceira seção serve como uma apresentação mais clara do que é o MAK, e para que ele é útil; a quarta parte é uma análise exploratória das variáveis que compõem o IDHM, e a apresentação dos resultados gerados pelo modelo; por fim, temos a conclusão, onde o trabalho como um todo será apurado.

2 REFERENCIAL TEÓRICO

2.1 O IDH E O IDHM

2.1.1 Índice de Desenvolvimento Humano (IDH)

O contexto de surgimento do IDH foi fortemente influenciado pela necessidade de suprir as deficiências dos Indicadores de Primeira Geração: o PIB e o PIB *per capita*. O problema destes indicadores era sua simplicidade e restrição, contemplavam apenas a produção econômica e a renda. Porém, conforme o entendimento que o crescimento econômico não provocava diretamente um aumento da qualidade de vida, iniciou-se a busca por novos indicadores capazes de quantificar o bem-estar social (GUIMARÃES & JANUZZI, 2005).

O IDH foi apresentado em 1990, no Relatório de Desenvolvimento Humano do Programa das Nações Unidas para o Desenvolvimento, foi idealizado pelo economista paquistanês Mahbub ul Haq, com a colaboração de Amartya Sen. O IDH é composto de três dimensões, que são centrais ao desenvolvimento social: as capacidades de se ter uma vida longa e saudável (saúde), acessar o conhecimento (educação), e possuir um padrão de vida digno (renda) (PNUD, 2013). Os indicadores que compõem o IDH são: esperança de vida à nascença (saúde), rendimento nacional bruto per capita (renda), média de anos de escolaridade (população com 25 anos ou mais) e anos de escolaridade esperados (educação). Esses três componentes são somados e é feita a medida dos três pra definir o IDH (PNUD, 2010).

O IDH teve um grande sucesso mundial, graças a sua simplicidade, a facilidade de sua compreensão, e pela sua forma abrangente de representar o desenvolvimento humano. Sintetizando em um único valor, as complexidades de três fatores centrais do bem-estar social. Além disso, serviu para fomentar discussões e reflexões sobre o significado do desenvolvimento humano para a sociedade (PNUD, 2013).

2.1.2 Índice de Desenvolvimento Humano Municipal (IDHM)

O IDHM é uma versão adaptada do IDH às realidades dos municípios brasileiros, desenvolvida em 1996, pela Fundação João Pinheiro, e pelo IPEA. Com o objetivo de estudar o nível de desenvolvimento humano dos municípios mineiros, no ano em questão. Apesar dos índices não serem comparáveis (o IDH e o IDHM), se referindo ao mesmo ano e local, ambos os índices aglomeram as mesmas realidades sociais: a saúde (longevidade), a renda e a educação. As principais diferenças entre o IDH e o IDHM, se dão nos indicadores de renda e educação, dado que neles se encontram as adaptações feitas, de modo a refletir com fidedignidade, o estado do desenvolvimento humano das populações residentes de cada município (GUIMARÃES & JANUZZI, 2005).

Existem diversos pontos positivos do uso do IDHM: o IDHM em seu enfoque de desenvolvimento nas pessoas, ao invés de crescimento econômico, populariza essa visão do desenvolvimento econômico, em contraponto ao PIB; quando ele sintetiza a complexa realidade das três dimensões levadas em conta, em um único valor, viabiliza a comparação entre os diversos municípios brasileiros ao longo do tempo; também estimula a melhoria das condições de vida nestes municípios, pois o ranking feito utilizando o IDHM, torna-se a base com que muitos formuladores e implementadores de políticas públicas no nível municipal agem, o que os impele a agir em prol de uma evolução das condições sociais dos cidadãos (PNUD, 2013).

Como exposto anteriormente, o IDHM é composto de três dimensões: saúde, renda e educação. Estes componentes do IDHM formam subíndices: IDHM Longevidade, IDHM Educação e IDHM Renda. Que são insumos de uma média geométrica, que os sintetiza em um único valor. O IDHM Longevidade, trata-se da expectativa de vida ao nascer; o IDHM Educação é composto de duas variáveis: a escolaridade da população adulta, e o fluxo escolar da população jovem (este indicador não compõe o IDH, sendo uma das adaptações próprias do IDHM); O IDHM Renda é formado apenas pela renda *per capita* do município (PNUD, 2013).

2.1.3 Crítica ao IDH e ao IDHM

Apesar dos inúmeros progressos, e vantagens, trazidos pelo IDH e o IDHM, eles não estão isentos de problemas. Um dos problemas citados por Guimarães e Januzzi (2005), abrange todos os indicadores sintéticos, e sua utilização pelos formuladores de políticas públicas, que é a substituição da métrica pelo conceito que ela representa. Logo, é esquecido o objetivo do desenvolvimento econômico e humano, que é substituído por uma obsessão pelo aumento do próprio índice. No caso do IDH e IDHM, o foco cego na métrica, pode eclipsar a realidade social, quando se inclui os outros problemas destes indicadores.

Outro problema semelhante dos indicadores sintéticos, é a crença de que sua existência, *per se*, garantiria uma melhor gestão dos recursos públicos e dos programas sociais. Mais uma crença que não colabora com a melhora do bem-estar social dos municípios, visto que os indicadores (como o IDHM), em seu uso indevido, não são bons critérios de elegibilidade de municípios para políticas sociais (GUIMARÃES & JANUZZI, 2005). Portanto, vale a pena retomar o problema simples porém pernicioso levantado por Veiga (2003): o método pelo qual o IDHM relaciona seus indicadores componentes (a média geométrica atualmente), não consegue identificar dois casos centrais de desenvolvimento humano, os locais de renda alta e condições sociais ruins, e os de baixa renda, porém com desenvolvimento social satisfatório.

O conjunto de problemas apresentado, e as conclusões tomadas por Guimarães e Januzzi (2005), revelam sérios problemas na tomada do IDHM como única forma de análise dos dados. Isto abre margem para uma inspeção mais ampla das condições sociais dos municípios, para o embasamento de políticas públicas direcionadas, apesar das inúmeras benesses do IDHM.

2.2 MAK E O DESENVOLVIMENTO ECONÔMICO E SOCIAL

Em 1982 o cientista da computação finlandês Teuvo Kohonen, concebeu um modelo de redes neurais conhecido como Mapa Auto-Organizável de Kohonen (MAK). O objetivo era a organização de grandes volumes de dados, de forma automática, porém preservando as relações topológicas entre os dados (KOHONEN, 1982).

Serão apresentados adiante os trabalhos que servem como inspiração a este, nos quais o MAK é utilizado em um contexto de análise do desenvolvimento econômico e social. São dois trabalhos, a monografia de Iven Martins Batista, e a dissertação de mestrado de Catarina Isabel Agostinho Cordas. Ambos trabalham o MAK de modo complementar aos indicadores sintéticos, de modo a produzir uma análise econômica e social mais abrangente.

O trabalho de Batista (2023) é a principal inspiração deste trabalho. O trabalho se propõe a abordar a análise de padrões socioeconômicos, ressaltando as limitações do IDH em capturar nuances e complexidades, como a desigualdade no padrão de vida (exemplificada pelo caso de Niterói). O MAK, sendo um método de aprendizado de máquina não supervisionado e uma Rede Neural Artificial (RNA), é utilizado para relacionar indicadores socioeconômicos (incluídos no cálculo do IDH) com o objetivo de captar um padrão de vida mais abrangente e inclusivo, englobando disparidades e desigualdades. O MAK mostra seu valor nas análises, ao por sua capacidade de realizar simultaneamente a representação gráfica e o agrupamento por *clusters* de observações baseado nas similaridades entre os dados, o que permitiu revelar grupos distintos de municípios no Rio de Janeiro.

O trabalho de Cordas (2011) apud Veiga (2003) utiliza o MAK como técnica de mineração de dados, visando lidar com a principal crítica em torno do IDH, a simplicidade com que suas dimensões são relacionadas pela média aritmética. Seu trabalho se baseia na literatura econômica que defende o entendimento de que o desenvolvimento humano é mais do que crescimento econômico, porém não deixa de criticar o IDH por deixar de fora dimensões importantes, como desigualdade, sustentabilidade e liberdade política. O objetivo principal é aplicar o MAK, que é um algoritmo apropriado ao problema devido às suas capacidades de agrupamento, na segmentação dos países do Relatório de Desenvolvimento Humano das Nações Unidas, utilizando mais dimensões de desenvolvimento. Os resultados obtidos com

o MAK sugerem que há diferenças na classificação dos países quando aplicadas dimensões adicionais, como *empowerment*, desigualdade e sustentabilidade, fornecendo uma contribuição de uma área recente (mineração de dados) para a análise do desenvolvimento humano.

3 METODOLOGIA

O objetivo do trabalho é a identificação de padrões no desenvolvimento socioeconômico entre os municípios mineiros utilizando os componentes do IDHM como variáveis, utilizando o Mapa Auto-Organizável de Kohonen. Pois, o MAK é capaz de agrupar municípios com características semelhantes, identificando relações subjacentes entre os dados, que se perdem ao sintetizar os componentes de Educação, Longevidade e Renda no IDHM. O MAK também preserva as relações topológicas existentes nos dados. (KOHONEN, 2001).

Antes de explorar o MAK, primeiro será introduzido o conceito de Rede Neural Artificial (RNA), pois o MAK é uma aplicação de redes neurais. As Redes Neurais Artificiais são algoritmos que tiveram sua criação inspirada nos neurônios do cérebro humano, e são compostas por unidades de processamento denominadas neurônios artificiais. Esses neurônios são organizados em camadas interconectadas capazes de processar informações e aprender padrões a partir dos dados fornecidos (HAYKIN, 2001). As RNAs têm sido amplamente utilizadas em tarefas de classificação, previsão, reconhecimento de padrões e mineração de dados (BISHOP, 1995).

De forma geral, uma RNA é composta por uma camada de entrada, responsável por receber os dados; uma ou mais camadas intermediárias, nas quais ocorre o processamento das informações; e por fim, uma camada de saída, responsável pela geração dos resultados. O aprendizado ocorre por meio do ajuste dos pesos sinápticos que conectam os neurônios da rede (HAYKIN, 1994).

Os Mapas Auto-Organizáveis de Kohonen fazem parte de uma categoria específica de RNA baseada em aprendizado não supervisionado. Diferentemente das redes supervisionadas, onde existe uma resposta desejada para guiar o treinamento, o aprendizado não supervisionado busca identificar padrões e estruturas presentes nos dados sem a necessidade de informações prévias sobre sua classificação (KOHONEN, 1982; KOHONEN, 2001).

O MAK foi desenvolvido por Teuvo Kohonen na década de 1980 com o objetivo de representar conjuntos de dados multidimensionais em espaços de menor dimensão, preservando suas relações topológicas. Em termos práticos, o MAK projeta observações complexas em um mapa bidimensional, permitindo ao mesmo tempo

redução de dimensionalidade e formação de clusters de observações com similaridades (KOHONEN, 2001).

No mapa gerado pelo modelo, municípios com perfis socioeconômicos próximos tendem a ser posicionados em regiões vizinhas do mapa, enquanto municípios com características distintas são alocados em regiões mais distantes (VESANTO; ALHONIEMI, 2000).

O MAK funciona através de duas etapas principais: competição e cooperação. Na etapa de competição, cada observação é ingerida pela rede como um vetor de entrada. Os neurônios do mapa calculam sua distância em relação a esse vetor e o neurônio que apresentar a menor distância é denominado Best Matching Unit (BMU), ou neurônio vencedor (KOHONEN, 2001).

A medida de similaridade utilizada neste trabalho é a distância euclidiana:

$$h(t) = \exp\left(\frac{-d^2}{2\sigma^2}\right)$$

onde:

- $d(x,y)$ representa a distância entre dois vetores;
- x_i corresponde ao valor observado da variável i ;
- y_i corresponde ao peso associado ao neurônio;
- n representa o número de variáveis utilizadas.

Após a identificação do neurônio vencedor, inicia-se a fase cooperativa. Nessa etapa, não apenas o BMU é atualizado, mas também os neurônios localizados em sua vizinhança topológica. O objetivo é fazer com que regiões próximas do mapa passem a representar observações semelhantes (KOHONEN, 2001).

A atualização dos pesos ocorre por meio de uma função de vizinhança, normalmente representada pela função gaussiana:

$$h(t) = \exp\left(\frac{-d^2}{2\sigma^2}\right)$$

em que:

- $h(t)$ representa a intensidade da influência do neurônio vencedor;
- d corresponde à distância entre neurônios no mapa;

- σ representa o raio de vizinhança.

Ao longo do treinamento, tanto a taxa de aprendizado quanto o raio de vizinhança diminuem gradualmente, permitindo que a rede passe de uma organização global para um refinamento local dos agrupamentos (BISHOP, 1995; HAYKIN, 1994).

É assim que o modelo funciona, agora pode-se seguir pra aplicação deste modelo no problema em questão, e avaliar os resultados gerados pelo MAK.

4 ANÁLISE DOS DADOS E RESULTADOS

4.1 ANÁLISE EXPLORATÓRIA DOS DADOS

4.1.1 Contextualização

Na formação econômica e política do Brasil, Minas Gerais foi um dos estados com grande destaque e importância. Principalmente durante períodos como o ciclo do ouro no século XVIII, no qual consolidou-se como principal polo minerador da colônia, com grandes contribuições para a dinamização econômica colonial e para a reorganização administrativa do território (BOXER, 2000; FURTADO, 2005). Minas Gerais também foi um importante centro agropecuário, siderúrgico e industrial, com capacidade de exercer influência política relevante no cenário nacional (BATELLA & DINIZ, 2006).

O estado, segundo o Censo Demográfico de 2010 (ano dos dados do trabalho), possuía cerca de 19,6 milhões de habitantes, sendo o segundo estado mais populoso do país, e também se configura como o estado com mais cidades, no total 853 (IBGE, 2010). O IDHM de Minas Gerais é enquadrado na categoria de Alto Desenvolvimento Humano, com o valor de 0,731, localizando-se um pouco acima da média nacional (PNUD; IPEA; FJP, 2013).

Para este trabalho, foram utilizados os indicadores socioeconômicos dos municípios mineiros referentes ao ano de 2010, componentes do Índice de Desenvolvimento Humano Municipal (IDHM). Os dados foram obtidos no Atlas do Desenvolvimento Humano no Brasil. Esta base foi escolhida pela amplitude dos dados que contemplam todas as cidades do estado, e oferece uma gama de variáveis envolvidas no cálculo do IDHM.

Na tabela 01 a seguir, serão expostas mais claramente as variáveis que serão utilizadas como input ao modelo, e as variáveis que serão utilizadas na comparação dos resultados:

Tabela 1 - Variáveis de interesse

Variável	Utilização
IDHM Longevidade	São as variáveis que serão utilizadas como input no MAK
IDHM Renda	
IDHM Educação	
IDHM Longevidade	As médias destas variáveis serão utilizadas na comparação dos clusters gerados pelo MAK
IDHM Renda	
IDHM Educação	
IDHM	

Fonte: Elaborada pelo autor.

4.1.2 Análise Descritiva

Para conseguir compreender os resultados gerados pelo modelo de clustering, é essencial a compreensão dos dados utilizados na alimentação do MAK. Com base nisso, será iniciada a etapa de análise exploratória dos dados. Começando com a exibição das medidas básicas das variáveis exibidas na tabela 02.

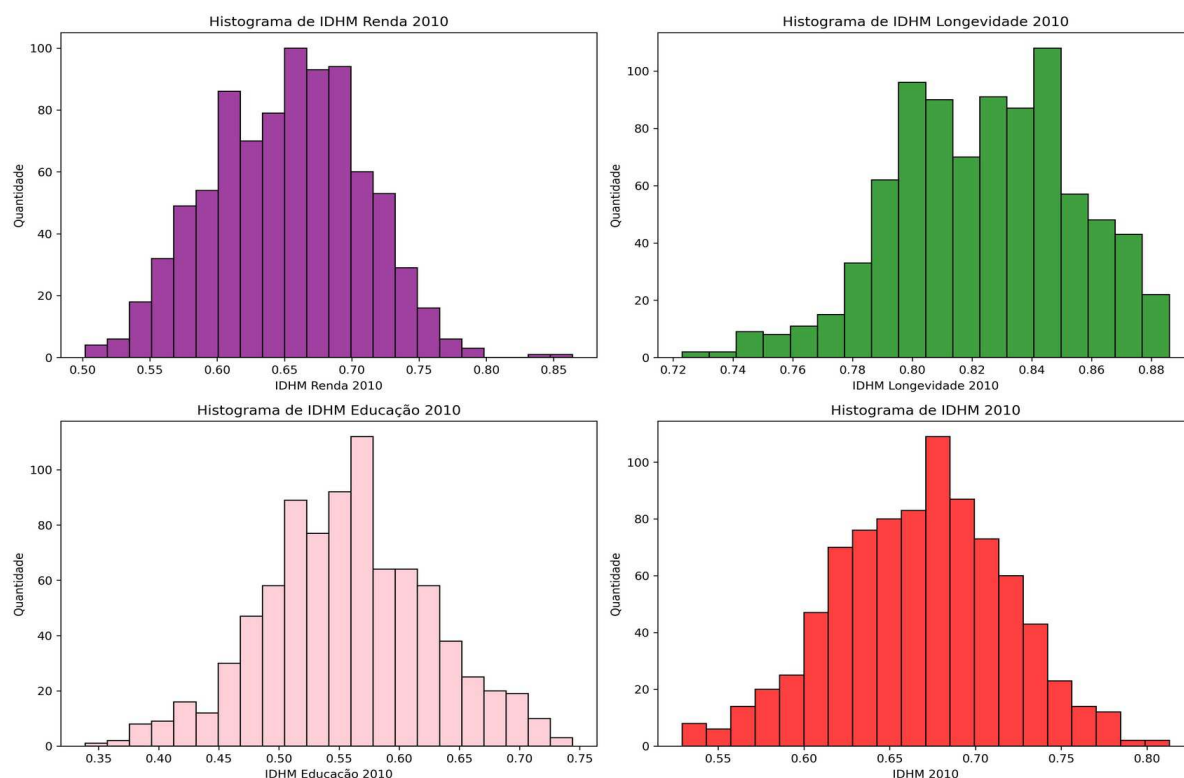
Tabela 2 - Medidas básicas das variáveis

Variáveis	Média	Desvio-padrão	Máximo	Mínimo
IDHM	0,667	0,049	0,813	0,529
IDHM Renda	0,652	0,055	0,864	0,502
IDHM Longevidade	0,823	0,029	0,886	0,723
IDHM Educação	0,556	0,069	0,744	0,339

Fonte: Elaborada pelo autor.

Apenas observando as medidas calculadas dos dados, já é possível identificar melhor a situação do desenvolvimento sócio-econômico de Minas Gerais. O Estado é classificado como de alto desenvolvimento, porém ao destrinchar o IDHM, nota-se que tanto a dimensão de renda, quanto à de educação, apresentam desempenho abaixo do esperado, sendo respectivamente médio e baixo. Enquanto o componente de saúde do IDHM se destaca, alcançando a classificação de muito alto. Também é perceptível a disparidade entre os valores máximos e mínimos no IDHM Renda e IDHM Renda, além de ambos apresentarem um desvio padrão maior que IDHM Longevidade. Todos os problemas citados quanto aos dois componentes, também se aplicam ao próprio IDHM das cidades, que na maioria das cidades está abaixo de alto. Um detalhe importante, é que a média de todos os IDHM das cidades, é diferente do IDHM do estado, devido a média geométrica do cálculo. As distribuições dos dados serão exibidas no gráfico 1, a seguir, como forma de revelar visualmente estas diferenças.

Gráfico 1 - Histogramas das variáveis analisadas

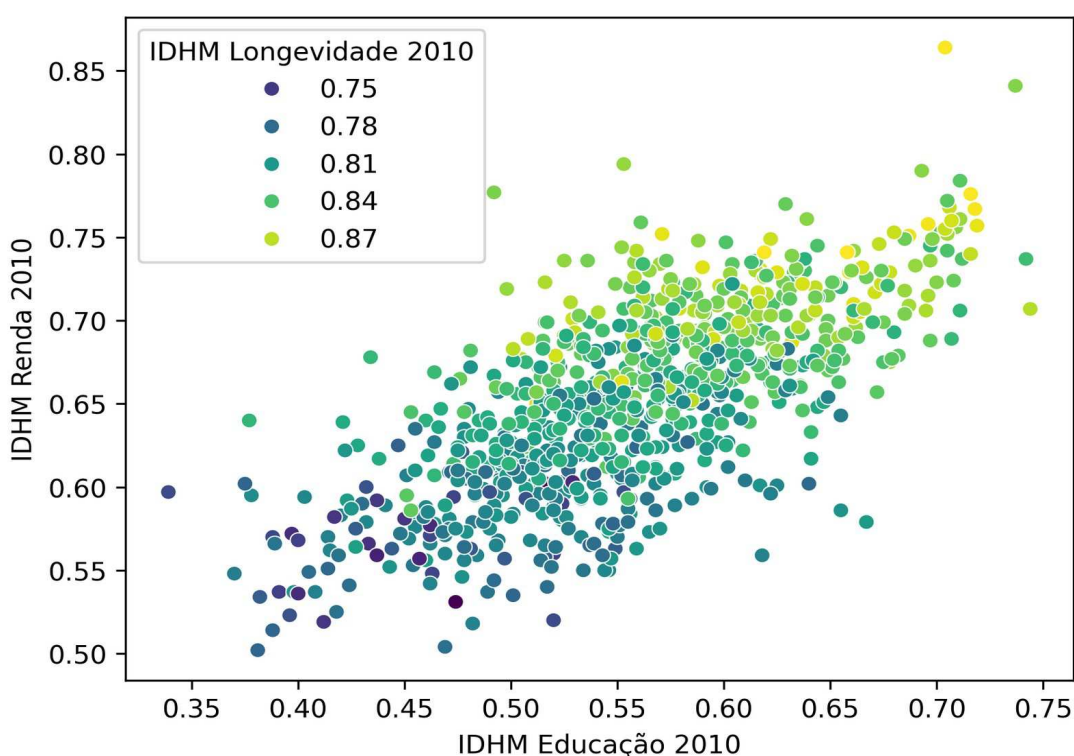


Fonte: Elaborado pelo autor.

Todas as distribuições de dados são semelhantes a uma distribuição normal. Portanto, as medidas sensíveis à não-normalidade, não foram afetadas. Observando os gráficos diretamente, fica muito clara a diferença do IDHM Longevidade quanto às outras variáveis, ela tem uma amplitude muito menor, e uma cauda direita mais larga, que é sinônimo de uma maior concentração dos dados, entre os valores mais altos próximos do máximo da variável.

No gráfico 2, a seguir, as variáveis componentes do IDHM, em suas dimensões de longevidade, educação e renda, serão relacionadas por meio de um gráfico de dispersão. O objetivo dessa visualização, é expor possíveis correlações entre os dados. E analisando o gráfico, observa-se a primeira vista, que tanto o IDHM Renda, quanto o IDHM Educação, aparentam estar correlacionados, pois o aumento de um é acompanhado pelo aumento do outro. E pela gradação das cores, o IDHM Longevidade parece ter algum grau de correlação com as outras duas variáveis.

Gráfico 2 - Dispersão com IDHM Educação, Renda e Longevidade



Fonte: Elaborado pelo autor.

As análises contidas nesta seção da monografia apenas expõem, através de gráficos e tabelas, informações básicas sobre as variáveis. Portanto, não se seguirão análises

mais aprofundadas, visto que o exibido é suficiente para uma melhor interpretação dos resultados do modelo.

4.2 RESULTADOS

É comum na literatura, e nas melhores práticas, utilizar a normalização como um método de pré-processamento dos dados, pois é comum que os dados tenham diferenças de escalas consideráveis, e isto seria um problema no momento de obter os resultados do modelo, pois poderiam gerar distorções da vizinhança topológica, afetando a qualidade dos resultados (BISHOP, 1995). Porém, como serão utilizadas medidas já compostas como o IDHM Renda, Educação e Longevidade, que possuem uma mesma escala, este tipo de tratamento de dados não será necessário.

Quanto à arquitetura da rede neural, foi decidido o uso de um mapa bidimensional de ordem 3x3. A escolha foi baseada nas dimensões dos dados de 3 colunas e 853 linhas. Portanto, para uma análise que identifique novas relações entre os dados, sem que ocorra uma fragmentação excessiva dos clusters, optou-se pelo mapa 3x3, em oposição ao trabalho de Batista (2023), que utiliza o 3x2; foram feitos diversos experimentos com diferentes mapas, e o 3x3 gerou os resultados mais interessantes no sentido de possibilitarem maior compreensão do problema em estudo.

Uma vez processados os dados, o modelo foi implementado na linguagem R, por meio do pacote “kohonen”. Foi escolhido, com base na literatura, o número de iterações de 500, valor utilizado nos trabalhos de referências desta monografia, que tratam de dados similares. (KOHONEN, 2001; BATISTA, 2023). Os outros parâmetros do modelo, a taxa de aprendizado e os pesos sinápticos, seguiram os valores padrão implícitos do pacote “kohonen”, que se mostraram adequados para a convergência do modelo.

Para a validação dos resultados, foram analisadas as médias e desvios padrões das variáveis agregadas por cluster gerado pelo MAK, em conjunto com a média do próprio IDHM. Esta forma de comparação permite a avaliação da medida em que os agrupamentos produzidos pelo modelo conseguem capturar novos padrões

no desenvolvimento sócio-econômico das cidades mineiras, não identificados pelo IDHM individualmente.

Os clusters gerados pelo MAK estão exibidos no gráfico 3, no formato de um mapa de calor, onde as cores mais quentes representam um maior número de dados, e as mais frias, um número menor. E as medidas dos cluster exibidas nas tabelas 3 e 4.

Tabela 3 - Médias dos clusters

Cluster	IDHM	IDHM Renda	IDHM Longevidade	IDHM Educação
1	0,762	0,743	0,861	0,694
2	0,700	0,708	0,852	0,570
3	0,664	0,662	0,831	0,532
4	0,723	0,702	0,851	0,634
5	0,683	0,655	0,819	0,594
6	0,625	0,623	0,813	0,484
7	0,640	0,603	0,801	0,544
8	0,599	0,567	0,787	0,484
9	0,564	0,564	0,779	0,410

Fonte: Elaborado pelo autor.

Tabela 4 - Desvio-padrão dos clusters

Cluster	IDHM	IDHM Renda	IDHM Longevidade	IDHM Educação
1	0,016	0,032	0,015	0,022
2	0,011	0,022	0,013	0,020
3	0,011	0,018	0,019	0,019
4	0,010	0,022	0,016	0,020
5	0,011	0,022	0,018	0,024
6	0,011	0,018	0,017	0,021
7	0,012	0,021	0,017	0,021
8	0,012	0,020	0,020	0,022
9	0,017	0,028	0,023	0,024

Fonte: Elaborado pelo autor.

Gráfico 3 - Mapa de calor dos clusters



Fonte: Elaborado pelo autor.

No cluster 1 reúnem-se as cidades de alto desenvolvimento. Este cluster apresenta o melhor desempenho geral, com média de IDHM de 0,763, IDHM Renda de 0,743, IDHM Educação de 0,694 e IDHM Longevidade de 0,861. Trata-se do grupo mais homogêneo (menores desvios-padrão). Concentra os principais centros urbanos e polos econômicos do estado, como Belo Horizonte, Uberlândia, Juiz de Fora, Poços de Caldas, Araxá, Barbacena, Timóteo e Viçosa. Em uma perspectiva regional, este grupo está fortemente associado à Região Metropolitana de Belo Horizonte, ao Triângulo Mineiro e a algumas cidades do Sul de

Minas, confirmando o padrão de desenvolvimento concentrado identificado por Batella e Diniz (2006).

O cluster 2 concentra as cidades em desenvolvimento médio-alto. Com média de IDHM de 0,701, destaca-se pela boa renda (IDHM Renda 0,708) e longevidade (IDHM Longevidade 0,852), ainda que a educação ainda seja apenas um pouco maior do que a média (IDHM Educação 0,571). Inclui municípios de médio porte com economia diversificada, como Alfenas, Canápolis, Guaranésia, Campina Verde e Cristais. Geograficamente, estes municípios situam-se predominantemente no Sul de Minas e em partes do Triângulo, representando uma camada intermediária de desenvolvimento regional. (BATELLA; DINIZ, 2006).

O cluster 3 reúne cidades de desenvolvimento médio, porém com uma saúde notável. Possui um IDHM médio de 0,664, com longevidade relativamente alta (IDHM Longevidade 0,831), mas educação um pouco abaixo da média (IDHM Educação 0,532), e um IDHM Renda na média de 0,662, representando uma renda mediana. Exemplos incluem Abre Campo, Aiuruoca, Belmiro Braga e Carmo de Minas. Este cluster revela uma das dissociações criticadas por Veiga (2003) e Guimarães e Januzzi (2005): boa condição de saúde e renda acima da média, mas uma educação que não acompanha os outros indicadores. Regionalmente, está presente em diversas áreas do interior central e da Zona da Mata.

O cluster 4 possui as melhores médias após o cluster 1, agrupando as cidades do segundo escalão mineiro. Com um IDHM médio de 0,723, com bom equilíbrio entre as dimensões (IDHM Renda 0,702; IDHM Educação 0,634; IDHM Longevidade 0,851). Exemplos de municípios: Andradas, Bambuí, Diamantina, Frutal, Curvelo e Alpinópolis compõem este grupo. Sua distribuição abrange regiões em expansão, especialmente no Sul, Centro e partes do Norte de Minas.

O cluster 5 é composto de cidades em médio desenvolvimento, porém com uma educação notável. A média de IDHM é 0,683, com longevidade alta (0,819), renda mediana (IDHM Renda 0,655) e uma educação acima da média (IDHM Educação 0,594). Exemplo de municípios: Aimorés, Alvinópolis, Andrelândia e Bom Jardim de Minas. Este perfil é comum em municípios de porte médio do interior, especialmente na Zona da Mata e em algumas áreas do Centro-Oeste mineiro.

O cluster 6 apresenta cidades de desenvolvimento médio-baixo com suas médias de IDHM de 0,625, IDHM Longevidade de 0,813 e IDHM Renda 0,623. Todos valores abaixo da média. Porém, sua característica marcante é o grande déficit

educacional presente nas cidades deste cluster, que possuem um IDHM Educação médio de 0,484, quase um desvio padrão abaixo da média de IDHM Educação das cidades mineiras, o segundo pior valor entre os clusters. Conta com forte presença de municípios da Zona da Mata, Vale do Rio Doce, Vale do Mucuri e áreas rurais do interior.

O cluster 7 aglomera as cidades em desenvolvimento médio-baixo. Se diferencia do cluster 6, com seu IDHM médio de 0,640, apesar de suas médias ruins em IDHM Renda (0,603) e IDHM Longevidade (0,801), valores com quase 1 desvio padrão abaixo da média, comparado ao cluster 6 ainda se sobressai pela grande diferença em IDHM Educação (0,544). Exemplo de municípios: Águas Formosas, Araçuaí, Arinos e Berilo. Este cluster está fortemente concentrado no Norte e Nordeste de Minas, especialmente no Vale do Jequitinhonha e Mucuri.

O cluster 8 é de baixo desenvolvimento, suas cidades têm um IDHM médio de 0,599, a educação é ruim (IDHM Educação 0,484), a renda também (IDHM Renda 0,567), e a saúde, apesar de ser considerada alta, está bem abaixo da média (IDHM Longevidade 0,787). Municípios como Angelândia, Berizal e Catas Altas da Noruega integram este grupo, predominantemente localizados em áreas periféricas do Norte de Minas.

O cluster 9 é composto pelos municípios com os piores níveis de desenvolvimento. Apresentam as piores médias de IDH (0,564), IDHM Educação (0,410), IDHM Longevidade (0,779), e IDHM Renda (0,564). A grande diferença entre o cluster 8 e o cluster 9 é sua educação em estado crítico. Este cluster concentra-se principalmente no Norte de Minas, Vale do Jequitinhonha e regiões do Rio Pardo. Novamente corroborando a literatura.

A distribuição espacial dos clusters gerados pelo MAK reproduz de forma clara o estado do desenvolvimento de Minas Gerais: os Clusters 1 a 4 concentram-se nas regiões mais dinâmicas (Metropolitana de Belo Horizonte, Triângulo Mineiro, Sul e Zona da Mata), enquanto os Clusters 6 a 9 predominam no Norte e Nordeste do estado. Os cluster gerados e suas características reforçam as análises de Batella e Diniz (2006) sobre a hierarquia urbana mineira e as limitações do IDHM em capturar tais desigualdades regionais (Guimarães & Januzzi, 2005; Veiga, 2003).

5 CONCLUSÃO

Este trabalho alcançou seu objetivo geral ao aplicar o Mapa Auto-Organizável de Kohonen aos componentes do IDHM dos municípios mineiros, gerando uma classificação alternativa e complementar ao IDHM. E também alcançou seus objetivos específicos nas etapas de: levantamento dos dados, preparação dos dados, aplicação do modelo e análise dos resultados.

A distribuição espacial dos clusters gerados pelo MAK reproduz de forma clara o estado do desenvolvimento de Minas Gerais: os Clusters 1 a 4 concentram-se nas regiões mais dinâmicas (Metropolitana de Belo Horizonte, Triângulo Mineiro, Sul e Zona da Mata), enquanto os Clusters 6 a 9 predominam no Norte e Nordeste do estado. Os clusters gerados e suas características reforçam as análises de Batella e Diniz (2006) sobre a hierarquia urbana mineira e as limitações do IDHM em capturar tais desigualdades regionais.

Ao contrário da média geométrica do IDHM, o Mapa Auto-Organizável de Kohonen permitiu identificar padrões multidimensionais e dissociações importantes entre as dimensões do desenvolvimento humano, oferecendo uma visão mais rica e territorializada, alinhada aos propósitos originais do índice.

A contribuição principal deste estudo reside em oferecer uma visão territorializada do desenvolvimento humano, capaz de subsidiar a formulação de políticas públicas mais focalizadas e eficientes, superando algumas limitações dos indicadores sintéticos tradicionais.

Como limitações do trabalho, destacam-se o uso de dados do Censo de 2010 (último ano com série completa) e a escolha subjetiva da arquitetura 3x3 do mapa.

Para futuros estudos, é interessante a incorporação de novas dimensões além dos componentes do IDHM, de modo a enriquecer ainda mais a análise; a utilização de dados mais recentes e a comparação com outros modelos de clusterização. Como por exemplo, adicionar novas variáveis que representem o desenvolvimento, desigualdade, infraestrutura, urbanização ou mercado de trabalho. Ou comparar os resultados do MAK com outros métodos de agrupamento, como k-means, clusterização hierárquica ou modelos de mistura.

Em resumo, o Mapa Auto-Organizável de Kohonen mostrou-se uma metodologia promissora e ainda pouco explorada na análise do IDHM no contexto

brasileiro. Sua aplicação em Minas Gerais reforça a importância de abordagens complementares aos indicadores oficiais, contribuindo para um entendimento mais profundo e com nuances geográficas do desenvolvimento humano no estado.

REFERÊNCIAS

BATELLA, Wagner B.; DINIZ, Alexandre M. A. **Desenvolvimento humano e hierarquia urbana: uma análise do IDH-M entre as cidades mineiras.** *Revista de Biologia e Ciências da Terra*, v. 6, n. 2, p. 367-374, 2006.

BATISTA, Iven Martins. **Análise comparativa do IDH do estado do Rio de Janeiro aplicando Mapas Auto-Organizáveis de Kohonen.** Juiz de Fora: Universidade Federal de Juiz de Fora, 2023.

BOXER, Charles R. **A Idade do Ouro do Brasil.** Rio de Janeiro: Nova Fronteira, 2000.

CORDAS, Catarina Isabel Agostinho. **Análise Exploratória do Índice de Desenvolvimento Humano: uma aplicação do Self Organizing Map na segmentação dos países.** [S.l.]: Universidade Nova de Lisboa, 2011.

FURTADO, Celso. **Formação Econômica do Brasil.** São Paulo: Companhia das Letras, 2005.

GUIMARÃES, José Ribeiro Soares; DE MARTINO JANNUZZI, Paulo. **IDH, indicadores sintéticos e suas aplicações em políticas públicas: uma análise crítica.** *Revista Brasileira de Estudos Urbanos e Regionais (RBEUR)*, v. 7, n. 1, p. 73-90, 2005.

HAYKIN, S. **Neural Networks: A Comprehensive Foundation.** New Jersey: Prentice Hall, 1994.

IBGE. **Censo Demográfico 2010.** Rio de Janeiro: IBGE, 2010.

KASKI, S.; KOHONEN, T. **Exploratory data analysis by the self-organizing map: structures of welfare and poverty in the world.** In: *Proceedings of the Third International Conference on Neural Networks in the Capital Markets*. World Scientific, 1996.

KAYANO, J.; CALDAS, E. L. **Indicadores para o diálogo.** São Paulo: Pólis; Programa Gestão Pública e Cidadania, EAESP/FGV, 2001.

KOHONEN, T. **Self-organized formation of topologically correct feature maps.** *Biological Cybernetics*, v. 43, p. 59–69, 1982.

KOHONEN, T. **Self-Organizing Maps.** Series in Information Sciences. Berlin: Springer-Verlag, 2001.

PROGRAMA DAS NAÇÕES UNIDAS PARA O DESENVOLVIMENTO. **Índice de Desenvolvimento Humano Municipal 2013.** Disponível em: <https://www.undp.org/pt/brazil/desenvolvimento-humano/publications/indice-de-desenvolvimento-humano-municipal-brasileiro-2013>.

PROGRAMA DAS NAÇÕES UNIDAS PARA O DESENVOLVIMENTO. **Relatório do Desenvolvimento Humano 2010.** Disponível em: <www.atlasbrasil.org.br>.

VEIGA, J. E. **Problemas do uso ingênuo do IDH-M.** *Valor*, 14 jan. 2003.

VESANTO, J.; ALHONIEMI, E. **Clustering of the Self-Organizing Map.** *IEEE Transactions on Neural Networks*, 2000.