

UNIVERSIDADE FEDERAL DE JUIZ DE FORA
FACULDADE DE ECONOMIA

IVEN MARTINS BATISTA

**ANÁLISE COMPARATIVA DO IDH NO ESTADO DO RIO DE JANEIRO
APLICANDO MAPAS AUTO-ORGANIZÁVEIS DE KOHONEN**

JUIZ DE FORA - MG
2023

IVEN MARTINS BATISTA

**ANÁLISE COMPARATIVA DO IDH NO ESTADO DO RIO DE JANEIRO
APLICANDO MAPAS AUTO-ORGANIZÁVEIS DE KOHONEN**

Monografia apresentada na Faculdade de
Economia da Universidade Federal de Juiz
de Fora, como requisito para obtenção do
título de bacharel em Ciências Econômicas.

Orientador: Prof. Alexandre Zanini

JUIZ DE FORA - MG
2023

Ficha catalográfica elaborada através do programa de geração automática da Biblioteca Universitária da UFJF, com os dados fornecidos pelo(a) autor(a)

Batista, Iven Martins.

Análise comparativa do IDH no estado do Rio De Janeiro aplicando Mapas Auto-Organizáveis de Kohonen / Iven Martins Batista. -- 2023.

31 p. : il.

Orientador: Alexandre Zanini

Trabalho de Conclusão de Curso (graduação) - Universidade Federal de Juiz de Fora, Faculdade de Economia, 2023.

1. Mapa de Kohonen. 2. IDH. 3. Indicadores sociais. I. Zanini, Alexandre, orient. II. Título.



UNIVERSIDADE FEDERAL DE JUIZ DE FORA
REITORIA - FACECON - Depto. de Economia

FACULDADE DE ECONOMIA / UFJF

ATA DE APROVAÇÃO DE MONOGRAFIA II (MONO B)

Na data de 05/12/2023, a Banca Examinadora, composta pelos professores

1 – Alexandre Zanini - orientador, e

2 – Rogério Silva de Mattos,

reuniu-se para avaliar a monografia do acadêmico IVEN MARTINS BATISTA, intitulada: **ANÁLISE COMPARATIVA DO IDH NO ESTADO DO RIO DE JANEIRO APLICANDO MAPAS AUTO-ORGANIZÁVEIS DE KOHONEN.**

Após primeira avaliação, resolveu a Banca sugerir alterações ao texto apresentado, conforme relatório sintetizado pelo orientador. A Banca, delegando ao orientador a observância das alterações propostas, resolveu **APROVAR** a referida monografia.

ASSINATURA ELETRÔNICA DOS PROFESSORES AVALIADORES



Documento assinado eletronicamente por Alexandre Zanini, Professor(a), em 05/12/2023, às 19:27, conforme horário oficial de Brasília, com fundamento no § 3º do art. 4º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



Documento assinado eletronicamente por Rogério Silva de Mattos, Professor(a), em 06/12/2023, às 08:54, conforme horário oficial de Brasília, com fundamento no § 3º do art. 4º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



A autenticidade deste documento pode ser conferida no Portal do SEI-Ufjf (www2.ufjf.br/SEI) através do ícone Conferência de Documentos, informando o código verificador 1612123 e o código CRC A5265C71.

RESUMO

O presente estudo visa abordar a análise de padrões socioeconômicos no estado do Rio de Janeiro por meio do uso do Mapa Auto-Organizável de Kohonen. O trabalho destaca a limitação do Índice de Desenvolvimento Humano (IDH) da Organização das Nações Unidas (ONU), propondo o MAK como uma ferramenta alternativa para capturar nuances e complexidades não consideradas pelo IDH. Os resultados revelaram clusters distintos entre si, agrupando municípios com características socioeconômicas similares dentro de cada grupo, assim como aspectos geográficos também.

Palavras chaves: Mapa de Kohonen; IDH; indicadores sociais.

ABSTRACT

The present study aims to address the analysis of socio economics standards in the state of Rio de Janeiro through the use of Self-Organising Map (SOM). The work highlights the limitations of the Human Development Index (HDI) from United Nations, proposing the SOM as a alternative tool to capture nuances not considered for the HDI. The results revealed distinct clusters, grouping municipalities with similar socioeconomic characteristics within each group, as well as geographical aspects.

Keywords: Self-Organising Map; HDI; social indexes.

LISTA DE SIGLAS

BMI – *Best Matching Unit*

IDH – Índice de Desenvolvimento Humano

MAK – Mapa Auto-Organizável de Kohonen

ONU – Organização das Nações Unidas

PNUD – Programa das Nações Unidas para o Desenvolvimento

RNA – Rede Neural Artificial

LISTA DE FIGURAS

FIGURA 1 - MODELO SIMPLES DE UMA RNA.....	17
FIGURA 2 – MAPA DE KOHONEN BIDIMENSIONAL.....	18
FIGURA 3 – RELAÇÃO DE VIZINHANÇA PELA FUNÇÃO GAUSSIANA	20
FIGURA 4 – MAK DE PROPORÇÃO BIDIMENSIONAL COM ORDEM 3X2.....	26
FIGURA 5 – REPRESENTAÇÃO NUMÉRICA DO MAK	26

LISTA DE GRÁFICOS

GRÁFICO 1 – RENDA PER CAPITA X IDH	23
GRÁFICO 2 – GRÁFICO DOS 5 MAIORES E 5 MENORES VALORES DO IDH	24
GRÁFICO 3 – RELAÇÃO DA TAXA DE INDIVÍDUOS COM ENSINO MÉDIO..	24
GRÁFICO 4 – RELAÇÃO DA RENDA POR NEURÔNIOS.....	27

LISTA DE TABELAS

TABELA 1 – VARIÁVEIS UTILIZADAS NO ESTUDO.....	22
TABELA 2 – RELAÇÃO DA MÉDIA E DESVIO-PADRÃO POR GRUPO.....	27
TABELA 3 – RELAÇÃO DOS MUNICÍPIOS POR CLUSTER.....	30

SUMÁRIO

1. INTRODUÇÃO	12
2. IDH E MAPA DE KOHONEN	14
2.1 IDH: CONCEPÇÃO, IMPORTÂNCIA E LIMITAÇÕES.....	14
2.2 RELAÇÃO DESENVOLVIMENTO HUMANO E MAPA DE KOHONEN	15
3. METODOLOGIA.....	17
4. ANÁLISE DOS DADOS E RESULTADOS	21
4.1 ANÁLISE EXPLORATÓRIA DOS DADOS	21
4.2 RESULTADOS	25
5. CONCLUSÃO.....	31
REFERÊNCIAS	32

1. INTRODUÇÃO

Ao longo do desenvolvimento da sociedade contemporânea, mais precisamente no século XX, teve-se a criação e desenvolvimento da Organização das Nações Unidas (ONU). Tal órgão tem o poder de discutir e participar ativamente em questões que envolvam conflitos geopolíticos entre os países, mudanças climáticas, programas sociais para os países de baixa renda, igualdade de direitos e oportunidades entre os indivíduos ao redor do mundo. Estas ações são desenvolvidas através de seu principal projeto, o Programa das Nações Unidas para o Desenvolvimento (PNUD) (MACHADO & PAMPLONA, 2008).

Dentre as formas para averiguar os melhores planejamentos para tomada de decisão, a ONU se baseia em diversos indicadores socioeconômicos, como renda per capita, taxa de analfabetismo, água encanada e esgoto tratado, expectativa de vida, concentração de renda, entre outros. Como forma de sintetizar tais indicadores e aprimorar essa mesma tomada de decisão, criou-se em 1990 o Índice de Desenvolvimento Humano (IDH), que tem o papel de juntar três dos principais indicadores socioeconômicos de uma localidade em apenas um, sendo eles (PNUD, 1990):

1. Longevidade (expectativa de vida desde o nascimento);
2. Conhecimento (taxa de analfabetismo é o principal indicador neste caso, porém é levada em conta também a escolaridade dos indivíduos ao longo da vida);
3. Padrão de vida (aqui temos 2 indicadores principais a serem considerados: PIB per capita e Paridade do Poder de Compra).

Todavia, o IDH é considerado por muitos um indicador de caráter genérico, que não consegue captar detalhes de desigualdade no padrão de vida dentro da área estudada, muitas das vezes sendo ocultos do objeto de análise (GUIMARÃES & JANNUZZI, 2005; MACHADO & PAMPLONA, 2008). Um bom exemplo, que também é pertinente para a realização deste trabalho, é o caso do município de Niterói, no estado do Rio de Janeiro. Mesmo possuindo o IDH mais alto dentre os municípios da região, é possível denotar uma

razão dos 10% mais ricos sobre os 40% mais pobres de 22,94, isto é, os ricos ganham cerca de vinte e três vezes mais que os pobres nestes percentis assinalados (RELATÓRIO DO DESENVOLVIMENTO HUMANO, 2010).

Com base nisto, este trabalho tem como objetivo relacionar os indicadores socioeconômicos para o cálculo do IDH com um método de aprendizado de máquina conhecido como Mapa Auto-Organizável de Kohonen. O objetivo é avaliar se é possível captar características de um padrão de vida mais abrangente do que o IDH proposto pela ONU, de forma mais inclusiva com variáveis que abarcam disparidades e desigualdades socioeconômicas.

O trabalho possui ao todo cinco seções, sendo elas: a introdução, já apresentada; uma seção detalhando o IDH e o Mapa de Kohonen; a metodologia, que abordará o método empregado com mais detalhes; uma seção de análise de dados e resultados, que mostra o que se pôde apurar com a execução do método mediante a base de dados e utilização do método; por fim, uma seção de conclusão, que evidencia os principais pontos levantados pelo estudo e a apuração do resultado.

2. IDH E MAPA DE KOHONEN

2.1 IDH: concepção, importância e limitações

O ser humano busca formas de avaliar seu bem-estar já há um tempo. Nesse intuito, faz-se presente a relevância da estatística como principal ferramenta de aferição deste elemento, que é bem representado no trabalho de Ignácio (2010). Porém, até 1990 tal ponderação era feita de forma simplista e primitiva, observando apenas variáveis macroeconômicas de forma isolada, como PIB, taxa de desemprego, inflação e etc. Com base nisto, teve-se a elaboração do Índice de Desenvolvimento Humano, originalmente criado por Mahbubul Haq em conjunto com o ganhador do Nobel de Economia Amartya Sen (UNDP, 1990). Originalmente a ideia era comparar o desenvolvimento humano entre países, tendo em conta as três variáveis já mencionadas na introdução para o cômputo e levando em consideração a paridade de poder de compra como principal variável comparativa.

O índice, em uma primeira análise, foi importante principalmente por se desprender das antigas unidades de medidas unilaterais que eram utilizadas para avaliar o desenvolvimento humano, que eram mais voltadas para mensuração da atividade econômica de um país (KAYANO & CALDAS, 2001). Conforme apontado por Guimarães e Jannuzi (2005), é um indicador que consegue ter um aspecto simples e que abrange os temas mais importantes referentes à qualidade de vida um indivíduo, além de ser de fácil comparação, já que a maioria das nações possui esses indicadores em suas bases de dados.

Entretanto, o índice ainda possui algumas falhas metodológicas consideradas graves em seu cálculo. Um trabalho notável que explicita uma dessas falhas é o de Sagar e Najam (1997), que critica a disparidade do fator renda entre os países, onde o PNUD utiliza-se de transformações logarítmicas para atenuar a diferença entre as nações ricas e pobres, mas que na verdade gera o efeito oposto. Outro fator importante é destacado no trabalho de Guimarães e Jannuzi (2005), onde estes consideram ser difícil comparar alfabetização e esperança de vida (indicadores de estoque), com PIB per capita e escolaridade bruta (indicadores de fluxo). Tal comparação pode acarretar em distorções intertemporais dos

resultados do IDH, pois estes indicadores possuem periodicidades diferentes na sua medição (PIB per capita, por exemplo, é medido anualmente enquanto que o levantamento da expectativa de vida acontece a cada 10 anos).

2.2 Relação desenvolvimento humano e Mapa de Kohonen

O Mapa Auto-Organizável de Kohonen (MAK) foi concebido originalmente em 1982 pelo cientista da computação Teuvo Kohonen. A ideia inicial era criar uma técnica que pudesse organizar automaticamente dados complexos em uma representação bidimensional, preservando as relações topológicas entre os dados originais (KOHONEN, 1982).

No que se refere a estudos sobre o envolvimento de desenvolvimento humano e mapas de Kohonen, já há uma literatura extensa. Um dos principais trabalhos neste segmento é refletido no artigo de Kaski e Kohonen (1996), que tem como objetivo analisar uma variedade de indicadores pertinentes ao bem-estar social e os sistemas econômicos de vários países através do MAK. Nele, os autores explicitam a importância da análise exploratória dos dados, onde o MAK realiza a tarefa de projeção (representação gráfica em mapa bidimensional) e agrupamento por clusters (grupos de observações com características similares entre si) simultaneamente, característica essa fundamental para escolha do método deste trabalho. Na seção de resultados, os autores constataram que o modelo conseguiu agrupar os exemplos apresentados de forma precisa, com a definição de seis grupos bem detalhados e próximos economicamente e socialmente, além de demonstrar também proximidade geográfica, embora não tenha sido apresentada nenhuma característica nesse sentido.

Outro trabalho importante a ser referenciado é o de Collan, Eklund e Back (2007), que agrupa países da Europa com características socioeconômicas similares sob a ótica de MAK, durante o período de 1998-2002. O objetivo do estudo é analisar 25 países de economias transitivas, com destaque principalmente para a Europa Oriental, que havia conquistado recentemente sua independência após o fim da União Soviética. Nas apurações, constataram-se seis clusters bem organizados e com padrões de desenvolvimento socioeconômicos distintos quando comparados uns aos outros. De

maneira geral, as novas nações independentes tiveram resultados bem similares umas as outras, com exceção da Estônia que disparou na maioria dos seus indicadores, e o Turqueministão apresentando os piores resultados.

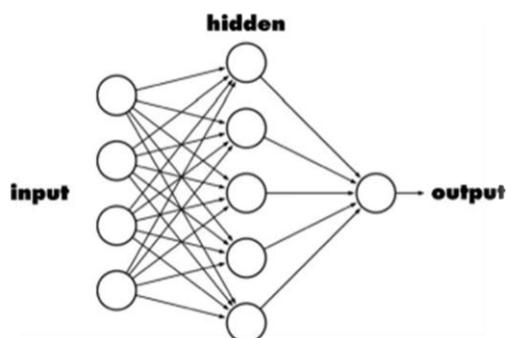
Em território brasileiro, vale ressaltar que é pouca ou quase inexistente a literatura abordando a aplicação de MAK com o IDH. Sendo assim, este trabalho monográfico utilizou-se como referência mais próxima o trabalho de Lins *et al* (2015), que analisa o IDH de municípios do Nordeste entre os anos 2000 e 2010. Seu objetivo é investigar a dependência espacial na evolução do índice durante o período assinalado, ou seja, detectar padrões espaciais de alto e baixo desenvolvimento e a relação de causalidade do espaço com este indicador, através de métodos da econometria espacial. Observando a análise exploratória dos dados, sendo a mais pertinente para o auxílio deste estudo, os autores concluíram que, de forma geral, municípios com alto IDH tendem a estarem mais próximos de outros municípios com um alto IDH, e o mesmo acontece para os valores mais baixos.

3. METODOLOGIA

Para se apreender os padrões de comportamento dos indicadores socioeconômicos e capturar uma relação com o IDH, este trabalho se baseou no método de Mapas Auto-Organizáveis de Kohonen. Na sequência, será feita uma breve introdução de redes neurais artificiais e em seguida, o método de MAK.

Antes de se introduzir o MAK devidamente, primeiro apresenta-se o conceito de rede neural artificial (RNA). De forma sucinta, uma RNA é um modelo computacional inspirado no funcionamento do cérebro humano, composto por neurônios artificiais interconectados que processam informações, aprendem com dados e são usados em uma variedade de tarefas, como classificação, previsão, reconhecimento de padrões e tomada de decisões (HAYKIN, 2001). Sua disposição se dá na seguinte forma: a) camada de entrada (*input*), onde são inseridos os exemplos da base de dados; b) uma camada escondida (*hidden layer*), que faz a intermediação entre entrada e saída da rede, além de cálculos durante o processo de aprendizado; c) camada de saída (*output*), que é o local onde se reflete os resultados dos cálculos. A forma como essas camadas se conectam e geram resultados é através de pesos sinápticos e uma função de ativação (HAYKIN, 2001), que serão detalhados dentro da apresentação do MAK. Um exemplo de como é uma RNA está representado na Figura 1.

Figura 1 - Modelo simples de uma RNA



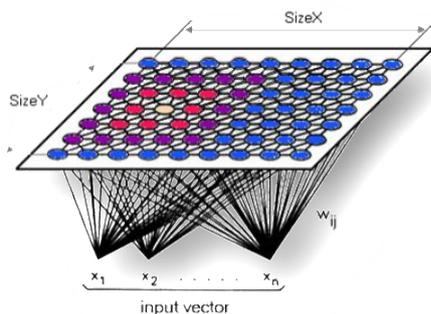
Fonte: Material didático (MATTOS, 2021)

A rede de MAK se enquadra como um método de rede neural com aprendizado não supervisionado, que se dá de forma competitiva e cooperativa (KOHONEN, 1982). Neste caso, o treinamento não é feito com a tutela de um “professor”, ou seja, não há uma saída desejada para os resultados, o que culmina em um aprendizado através da competição e cooperação entre neurônios, e não com os erros, como é no caso do treinamento supervisionado (HAYKIN, 2001).

Sendo assim, o MAK visa construir um mapa topológico artificial, se baseando na auto-organização dos neurônios e resultando em um mapa discreto com dimensão arbitrária, em um aprendizado competitivo/cooperativo (KOHONEN, 1982). É composto por duas camadas: uma de entrada, onde são apresentados os dados juntamente com os pesos, e uma de saída, que é onde se indica a dimensionalidade do mapa (neste caso, iremos tratar de um mapa bidimensional) (KOHONEN, 2001).

A rede é totalmente interconectada, isto é, a camada de entrada se conecta com todos os neurônios da camada de saída, e estes por sua vez são interligados entre si. A ideia é que os neurônios da camada de saída respondam a um determinado sinal sináptico emitido pela camada de entrada e, desta forma, se agrupem com outros que possuam o mesmo sinal sináptico (KOHONEN, 2001). Uma melhor visualização desse processo poder ser compreendida na Figura 2.

Figura 2 – Mapa de Kohonen Bidimensional



Fonte: (BASTOS & MARTINI, 2013)

Conforma mostra a imagem, têm-se a camada de entrada formada pelo vetor (x_1, x_2, \dots, x_n) , os pesos que são computados em W_{ij} , e a camada de saída forma pelo plano (X, Y) (BASTOS & MARTINI, 2013).

O método consiste primeiramente na competição de neurônios, chamado de *Best Matching Unit* (BMU), onde o neurônio que apresentar um “melhor casamento”, isto é, a menor distância da camada de entrada – uma vez calculada todos os pesos de conexão – com a camada de saída é ativado (KOHONEN, 2001). Tal processo pode ser medido através da distância euclidiana, evidenciada da seguinte forma:

$$D(x,y) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2}$$

Onde:

- $D(x,y)$ é a distância entre os vetores x e y .
- n é o número de características no vetor.
- x_i e y_i são as características dos vetores x e y .

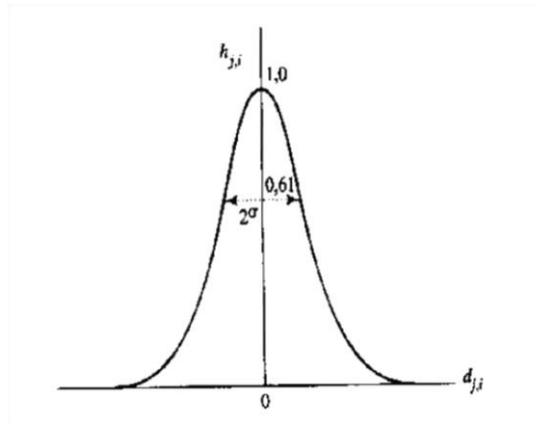
Finalizada esta etapa, acontece a fase de aprendizado, onde o neurônio ativado atualiza o seu vetor de pesos e o dos neurônios da sua vizinhança topológica, de forma a se assemelhar com os dados da camada de entrada (KOHONEN, 2001). Esta atualização é tipicamente feita com uma função de vizinhança gaussiana (exemplificada graficamente na Figura 3), dada pela seguinte equação:

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

Nesta equação, temos a seguinte relação:

- $h(t)$ é a função de vizinhança.
- d é a distância entre o neurônio vencedor e o neurônio atual.
- σ é o raio da vizinhança, que diminui com o tempo para permitir uma redução gradual da vizinhança.

Figura 3 – Relação de vizinhança pela função gaussiana



Fonte: (BISHOP, 1995)

Uma vez que um neurônio é ativado, a atualização se dá similar às cores na Figura 2, isto é, aqueles mais próximos do ativo (bege), possuem uma atualização maior do peso (vermelhos), se comparado aos vizinhos mais distantes (roxos). Este processo de competição e cooperação pode ser repetido em um número específico de épocas (iterações), ou até a convergência da rede, isto é, até os neurônios se organizarem em *clusters* que representem adequadamente a estrutura dos dados de entrada (KOHONEN, 2001).

Depois de evidenciado o modelo, pode-se seguir para a etapa de análise dos dados e resultados, a fim de averiguar a apuração do método e comparar se é de fato eficiente para este caso.

4. ANÁLISE DOS DADOS E RESULTADOS

4.1 Análise exploratória dos dados

O estado do Rio de Janeiro desempenhou um papel político e econômico (e ainda desempenha na questão econômica) de suma importância para o Brasil. Este é considerado fundamental para as atividades agroexportadoras e siderúrgicas ao longo da história, além de abrigar a capital nacional por mais de um século, que trouxe investimentos sólidos e significativos para o estado no que concerne à infra-estrutura, cultura, educação e saúde (AZEVEDO, 2011). Conforme os dados do Censo de 2010, o Rio de Janeiro apresentava uma população de aproximadamente 16 milhões de habitantes, renda per capita de 666 reais e um IDH de 0,709 (RELATÓRIO DO DESENVOLVIMENTO HUMANO, 2010). Logo, mostrou-se um candidato pertinente como alvo para a realização deste estudo.

Para a construção deste trabalho, foram-se escolhidos indicadores socioeconômicos dos municípios do estado do Rio de Janeiro referentes ao ano de 2010. Todos os dados tiveram como fonte o Censo de 2010, obtido através do Relatório do Desenvolvimento Humano (2010). Escolheu-se essa base por ser a única com todos os indicadores necessários para análise, de acordo com a amostra selecionada.

Seguindo a literatura (ONU, 1990), foram escolhidas oito variáveis para o cálculo do MAK, e quatro para fins comparativos, que estão dispostas na Tabela 1:

Tabela 1 – Variáveis utilizadas no estudo

Variável	Objetivo
Taxa de atividade - 18 anos ou mais de idade	Introduzidas no cálculo (input)
Renda per capita	
Porcentagem de domicílios com banheiro e água encanada	
Expectativa de vida	
Mortalidade infantil	
Taxa de analfabetismo – 25 anos ou mais de idade	
Porcentagem de indivíduos com ensino médio completo – 25 anos ou mais de idade	
Porcentagem de indivíduos com ensino superior completo – 25 anos ou mais de idade	
Coeficiente de Gini	
Razão dos 10% mais ricos sobre os 40% mais pobres	
IDHM (Índice de Desenvolvimento Humano Municipal)	
População	

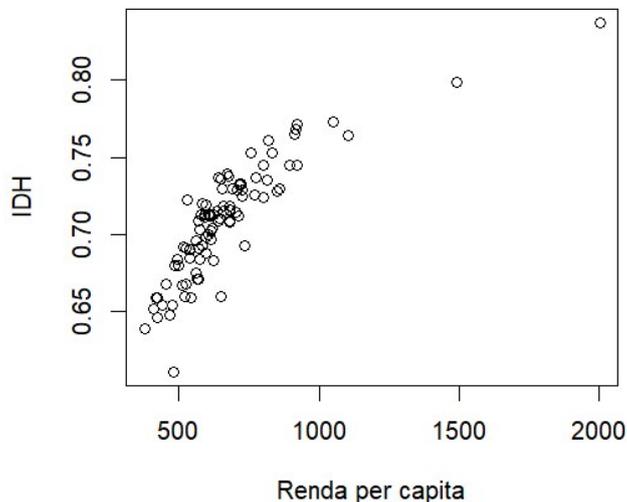
Nota: Todas as variáveis foram extraídas do Censo Demográfico de 2010

Fonte: Elaborado pelo autor

Analisando as variáveis, ponderou-se necessário fazer observações exploratórias sobre as mesmas. No Gráfico 1, por exemplo, foi elaborado um gráfico que demonstra a relação de dispersão entre o IDH dos municípios com suas respectivas rendas per capita. É interessante notar que apenas 35 municípios (38% do total) possuem uma renda per capita acima da média, o que já pode deixar alguns indícios de concentração de renda em determinadas localidades.

Gráfico 1 – Renda per capita X IDH

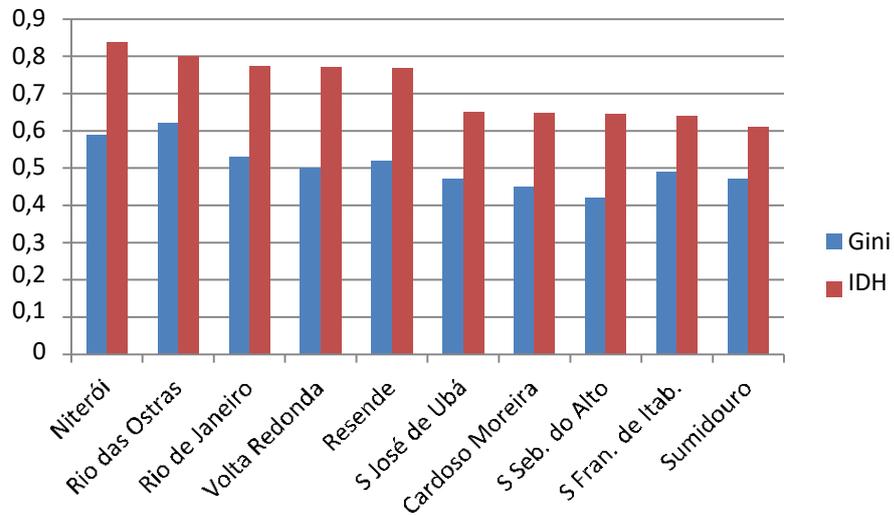
Dispersão Renda per capita x IDH



Fonte: Elaborado pelo autor

Outro dado interessante a se observar é a comparação dos cinco municípios com os maiores IDH em relação aos cinco menores, representado no Gráfico 2. Mesmo se tratando de um estado pequeno em número de municípios e em território, é perceptível a diferença dos valores e da construção dos parâmetros socioeconômicos ao longo do estado, o que demonstra a versatilidade das atividades econômicas do mesmo. Junto a isso, optou-se por inserir também o índice de Gini para cada um dos municípios. É notável como os municípios de menores IDH possuem também os menores valores para o coeficiente de Gini, que pode sinalizar para este caso um nivelamento na pobreza destes municípios, pois estes também apresentam os menores indicadores econômicos para o estado.

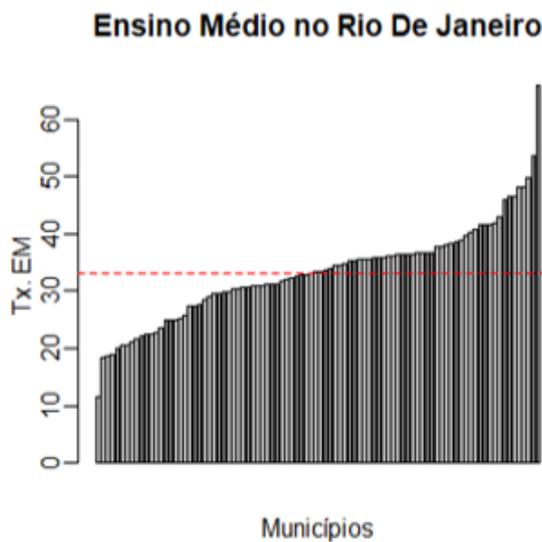
Gráfico 2 – Gráfico dos 5 Maiores e 5 Menores Valores do IDH



Fonte: Elaborado pelo autor

Uma variável que chamou a atenção foi a disparidade entre os municípios para a taxa de ensino médio completo. Só em termos de amplitude, a menor taxa (Sumidouro) vale por quase 1/6 da maior (Niterói). No Gráfico 3, pode-se observar a distribuição dessa variável ao redor do estado de forma mais nítida, juntamente com a média também atribuída no gráfico, configurada pelo tracejado vermelho.

Gráfico 3 – Relação da taxa de indivíduos com ensino médio



Fonte: Elaborado pelo autor

4.2 Resultados

Antes de realizar o input dos dados de entrada, foi necessária a operação de normalização como método de pré-processamento, visto que os indicadores quando comparados uns aos outros possuem uma diferença de escala considerável. Isto acarretaria em uma distorção da vizinhança topológica, dificultando a obtenção de um resultado fidedigno. Seguindo a literatura, optou-se por transformar as variáveis linearmente para ter média zero e desvio padrão igual a um (BISHOP, 1995).

Com os dados normalizados, seguiu-se para a etapa de modelação do MAK em si. Com relação à arquitetura da rede, elaborou-se um mapa bidimensional de ordem 3x2. A justificativa para esta grandeza se dá pelo tamanho da base de dados (92 observações) e pelo grau de complexidade das variáveis, além de serem poucas, o que permite uma análise mais rica em detalhes.

Possuindo todas as informações processadas, o mapa foi gerado utilizando o software R, através principalmente de 2 bibliotecas: “RSNNS” e “NeuralNetTools”. O número de iterações estipulado foi 500, baseado no trabalho de Cortés, Martínez e Rubio (2003), que utilizaram variáveis similares tanto em quantidade quanto em características, assim como um número semelhante de observações. Os valores para taxa de aprendizado (definido na função de vizinhança gaussiana) e dos pesos sinápticos estão implícitos por *default*. Tais valores estão dentro do pacote “RSNNS”, que atua de maneira própria para a convergência dos resultados. Já o pacote “NeuralNetTools” foi utilizado para representar a visualização dos resultados. Como métrica de validação dos resultados, compararam-se os clusters obtidos com o IDH de cada município.

A visualização do mapa está representada na Figura 4, onde as cores quentes representam um cluster com um maior número de municípios, enquanto que as frias representam o oposto. Além disso, enumeraram-se cada um desses grupos, que serão detalhados na sequência. A Figura 5 demonstra, por sua vez, a representação numérica da quantidade de municípios em cada cluster.

Figura 4 – MAK de proporção bidimensional com ordem 3x2



Fonte: Elaborado pelo autor

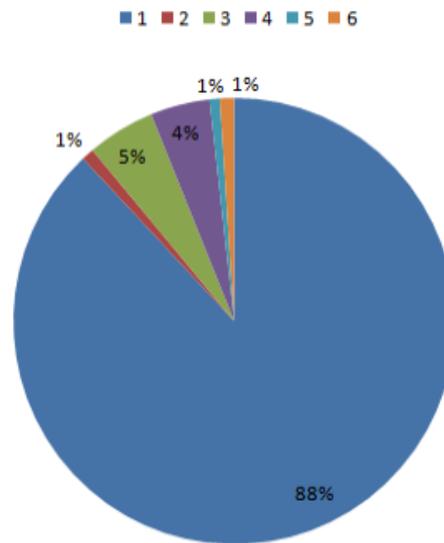
Figura 5 – Representação numérica do MAK

30	10
25	13
5	9

Fonte: Elaborado pelo autor

Comparando inicialmente os clusters definidos nos neurônios do modelo, há algumas características consideráveis para reconhecimento. A começar pela renda bruta (oriunda da renda per capita e população dos municípios), há uma elevada concentração da mesma no primeiro neurônio (88% da renda do estado), e que comporta apenas 30 municípios. Os 12% restante da renda estão distribuídos de forma irregular nos outros neurônios. O Gráfico 4 elucidava melhor essa disposição. Na Tabela 2, está disposto o resultado dos grupos, com a média e o desvio-padrão das principais variáveis utilizadas no cômputo do método.

Gráfico 4 – Relação da renda por neurônios



Fonte: Elaborado pelo autor

Tabela 2 – Relação da média e desvio-padrão por grupo

Média								
Neurônios	Rpc	TxAt	Rrp	Analf.	E.M.	Mi	EV	IDH
1	841,42	66,05%	14,21	4,99%	40,01%	13,72	75,25	0,742
2	602,21	63,45%	13,10	9,15%	32,04%	15,71	73,84	0,706
3	517,49	62,86%	11,22	11,26%	25,24%	16,98	73,00	0,672
4	677,60	61,86%	12,17	5,37%	38,39%	16,24	73,48	0,724
5	620,17	69,07%	10,05	7,14%	34,58%	16,16	73,54	0,711
6	573,54	68,25%	11,22	9,67%	25,96%	16,64	73,37	0,683

Desvio-Padrão								
Neurônios	Rpc	TxAt	Rrp	Analf.	E.M.	Mi	EV	IDH
1	288,50	3,37%	3,74	1,48%	7,43%	1,05	0,62	0,029
2	58,16	2,55%	2,23	1,71%	3,59%	0,70	0,48	0,017
3	74,78	4,53%	2,65	3,44%	4,92%	0,92	0,57	0,023
4	62,31	3,88%	1,71	1,41%	3,97%	0,65	0,42	0,013
5	88,32	3,54%	1,69	1,51%	1,92%	0,42	0,29	0,013
6	77,72	2,26%	1,63	1,57%	5,72%	0,98	0,71	0,024

Nota: Rpc = renda per capita; TxAt = taxa de atividade; Rrp = razão dos 10% mais ricos sobre os 40% mais pobres; Analf. = taxa de analfabetismo; E.M. = taxa de indivíduos com ensino médio completo; Mi = mortalidade infantil; EV = expectativa de vida; IDH = Índice de Desenvolvimento Humano.

Fonte: Elaborado pelo autor

Tomando uma análise mais minuciosa, os municípios localizados no neurônio 1 são os mais ricos do estado. Possuem os valores mais altos para a maioria dos indicadores, com destaque para a população (13 milhões) e renda (12,2 bilhões). Os municípios participantes são, em sua maioria, da região metropolitana do estado, denotando a forte influência espacial desse grupo, com a presença da capital que concentra grandes investimentos em uma ampla gama de atividades econômicas, tendo como carro-chefe a exploração de petróleo e gás, atraindo mão de obra e capital dos municípios vizinhos. Porém, este mesmo grupo possui um desvio-padrão consideravelmente alto para algumas variáveis, com destaque para renda per capita, razão dos 10% mais ricos para os 40% mais pobres e mortalidade infantil. Logo, pode-se concluir que há uma elevada discrepância dos municípios mais ricos para os mais pobres dentro deste mesmo grupo, influenciada principalmente pela capital, o que mais uma vez enaltece a altíssima concentração das atividades econômicas em poucos municípios no estado.

O neurônio 2 possui 10 municípios, que totalizam 1% da renda do estado. Geograficamente, é distribuído ao longo do estado, desde a região do Centro-Sul fluminense até a região Norte, não possuindo uma relevância espacial significativa. Com relação as suas variáveis, são mais estáveis e consistentes do que o grupo 1, porém ainda com uma concentração de renda significativa (Rrp), ocupando a 2ª posição dentre os 6 grupos. Seu principal município é Bom Jesus do Itabapoana, que abriga 20% (25 milhões de reais) da renda e 17% (35 mil habitantes) da população dentro deste grupo.

O terceiro grupo detém 25 municípios, e é considerado o mais pobre dentre os 6, tendo a grande maioria dos indicadores mais baixos que os outros. Possui 5% da renda de todo o estado, todavia com uma alta concentração no município de Belford Roxo na região metropolitana. É um grupo bem disperso ao redor do território fluminense, porém com uma nítida visualização da marginalização dos seus integrantes. Tal fato pode ser explicado por suas características geográficas, visto que não são consideradas próximas de nenhum grande polo industrial de maneira geral, caracterizando o que seria o interior do estado.

O cluster 4, que possui 13 municípios por sua vez, tem uma característica que pode ser considerada única. Entre os 6 grupos, ele é o mais “consistente”, apresentando apenas

municípios com um alto IDH, seguindo a classificação dada pelo PNUD (2013). Possui o menor desvio-padrão do IDH, este juntamente com o grupo 5, além da taxa de analfabetismo. O aspecto geoeconômico neste cluster é o ponto de destaque, pois 9 municípios estão na região Centro-Sul e Médio Paraíba (sendo estes bem próximos uns dos outros) e os outros 3 na Baixada Litorânea. Tal fato pode ser explicado principalmente pelo município de Três Rios no Centro-Sul, que exerce uma função chave de logística, ligando as rodovias do Centro-Sul com o Médio Paraíba.

O neurônio 5 é o menor com apenas 5 municípios. Além do menor desvio-padrão para o IDH conforme já citado, este grupo tem como destaque a mais alta taxa de atividade (69,07%). Adicionalmente, o grupo apresenta a menor taxa da razão rico sobre pobres, onde os 10% mais ricos ganham, em média, dez vezes mais que os 40% mais pobres. O principal município neste grupo é Casimiro de Abreu, com uma renda per capita de 768 reais e uma taxa de indivíduos com ensino médio de 36,44%.

Por último, mas não menos importante, é citado o grupo 6. Este se assemelha muito com o grupo 2, por ter seus municípios bem dispersos ao redor do território fluminense. Com relação aos seus indicadores socioeconômicos, são considerados os mais baixos em sua maioria, tendo inclusive a menor taxa de indivíduos com ensino médio completo (25,96%), caracterizando-o como o segundo grupo mais pobre do estado. Dos seus 9 municípios, 6 estão na faixa de IDH médio, e 3 na faixa de IDH alto.

A disposição dos municípios por grupos está evidenciada na Tabela 3 logo abaixo:

Tabela 3 – Relação dos municípios por cluster

Neurônios	Municípios
1	Angra dos Reis, Araruama, Armação dos Búzios, Cabo Frio, Campos dos Goytacazes, Cordeiro, Duque de Caxias, Itatiaia, Iguaba Grande, Itaperuna, Itaguaí, Macaé, Volta Redonda, Valença, Magé, Mangaratiba, Maricá, Mesquita, Miguel Pereira, Niterói, Nova Friburgo, Nova Iguaçu, Paraty, Petrópolis, Resende, Rio de Janeiro, Rio das Ostras, S Gonçalo, S João de Meriti, Teresópolis
2	Areal, Bom Jesus do Itabapoana, Carmo, Itaocara, Macuco, Miracema, Natividade, Paraíba do Sul, Quissamã, Rio das Flores
3	Aperibé, Belford Roxo, Cambuci, Carapebus, Cardoso Moreira, Itaboraí, Italva, Japeri, Laje do Muriaé, Paty do Alferes, Porciúncula, Queimados, Rio Bonito, Rio Claro, Santa Maria Madalena, S Fidélis, S Fran. de Itab., S João da Barra, S José de Ubá, S Seb. do Alto, Sapucaia, Silva Jardim, Sumidouro, Tanguá, Trajano de Moraes
4	Arraial do Cabo, Barra do Pirai, Barra Mansa, Engenheiro Paulo de Frontin, Mendes, Nilópolis, Paracambi, Pinheiral, Pirai, S Pedro da Aldeia, Saquarema, Três Rios, Vassouras
5	Casimiro de Abreu, Conceição de Macabu, Porto Real, Quatis, Seropédica
6	Bom Jardim, Cachoeiras de Macacu, Cantagalo, Comendador Levy Gasparian, Duas Barras, Guapimirim, Santo Antônio de Pádua, S José do Vale do Rio Preto, Varre-Sai

Fonte: Elaborado pelo autor

5. CONCLUSÃO

Este trabalho visou apresentar, através do aprendizado de máquina com o método de Mapas Auto-Organizáveis de Kohonen, padrões de comportamento para indicadores socioeconômicos dos municípios do estado do Rio De Janeiro em comparação ao IDH dos mesmos. Para fins de conferência do método, utilizou-se do próprio IDH desses municípios como forma de validação do mapa.

De forma geral, obtiveram-se bons resultados e que foram congruentes com a teoria econômica, principalmente em 3 dos 6 grupos, atingindo o objetivo do trabalho. Quando comparado ao IDH, demonstrou-se uma boa flexibilidade e junção de municípios com características similares, com ênfase na perspectiva econômica em relação às educacionais e sanitárias, além de agregar localidades semelhantes no aspecto geoeconômico.

Há várias sugestões para aprimorar os resultados aqui encontrados. Uma delas se dá no próprio método utilizado, que seria através da mudança de *software*, ou nas bibliotecas utilizadas, de maneira que possa se manipular as taxas de aprendizados e os pesos sinápticos. Isto poderia acarretar em uma convergência mais robusta dos resultados, configurando um rearranjo dos grupos. Além disto, inserir outras variáveis no cálculo, como por exemplo, a porcentagem de indivíduos em situação de vulnerabilidade ou razão de vulnerabilidade social, o que traria resultados mais minuciosos sobre estas perspectivas, uma vez que a desigualdade social no estado como um todo é uma característica muito forte a ser tratada.

REFERÊNCIAS

AZEVEDO, Aline Oliveira. **Análise do IDH do estado do Rio de Janeiro**. Trabalho de Conclusão de Curso (Bacharelado em Ciências Econômicas) – Instituto Três Rios, Departamento de Ciências Econômicas e Exatas, Universidade Federal Rural do Rio de Janeiro, Três Rios, 2011.

BISHOP, C.M.(1995): **Neural Networks for Pattern Recognition**. Ed. Clarendon Press, Oxford.

COLLAN, Mikael; EKLUND, Tomas; BACK, Barbro. **Using the Self-Organizing Map to Visualize and Explore Socio-Economic Development**. EBS Review 22(1), 6-15, 2007.

CORTÉS, Esteban Alfaro; MARTÍNEZ, Matías Gámez; RUBIO, Noelia García. **Una clasificación socioeconómica de las regiones europeas mediante mapas de Kohonen**. Universidad de Castilla-La Mancha, Facultad de Ciencias Económicas y Empresariales, 2003.

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. (1994): **Neural Networks. A Comprehensive Foundation**. Ed. Prentice Hall, New Jersey.

IGNÁCIO, Sérgio Aparecido. **Importância da estatística para o processo de conhecimento e tomada de decisão**. Revista Paranaense de Desenvolvimento 118 (2010): 175-192.

KASKI, S; KOHONEN, T. **Exploratory data analysis by the self-organizing map: structures of welfare and poverty in the world**. Proceeding 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. (1982): **Self-organized formation of topologically correct feature maps**. Biological Cybernetics 43:59 - 69.

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

LINS, J. G. M. G. et al. **Análise espacial da evolução do índice de desenvolvimento humano nos municípios da região nordeste**. Revista Economia e Desenvolvimento, v. 14, n. 1, p. 81-96, 2015.

M. R. Bastos and I. S. C. Martini, **An approach free model for evaluation of electricpower systems with the use of artificial intelligence**. 2013 IEEE PES Conference

on Innovative Smart Grid Technologies (ISGT Latin America), Sao Paulo, Brazil, 2013, pp. 1-4, doi: 10.1109/ISGT-LA.2013.6554497.

MACHADO, João Guilherme Rocha; PAMPLONA, João Batista. **A ONU e o desenvolvimento econômico: uma interpretação das bases teóricas da atuação do PNUD**. Economia e sociedade, v. 17, p. 53-84, 2008.

MATTOS, Rogério. **Aprendizado de máquina para economia**. Material Didático, 2021

PINTO, Daniela Gomes Coordenação; COSTA, Marco Aurélio Coordenação; MARQUES, Maria Luiza de Aguiar Coordenação. **O índice de desenvolvimento humano municipal brasileiro**. Brasília: Programa das Nações Unidas para o Desenvolvimento, Instituto de Pesquisa Econômica Aplicada, 2013.

PROGRAMA DAS NAÇÕES UNIDAS PARA O DESENVOLVIMENTO. **Relatório do desenvolvimento humano 2010**. Disponível em: <www.atlasbrasil.org.br>. Acesso em: 23 out 2023.

SAGAR, Ambuj D.; NAJAM, Adil. **The human development index: a critical review**. *Ecological economics*, v. 25, n. 3, p. 249-264, 1998.

UNDP. **Human development report**. Oxford University Press, 1990.