

ANÁLISE DA COMPOSIÇÃO DE ÍNDICES DE DESENVOLVIMENTO UTILIZANDO MINERAÇÃO DE DADOS

Samira de Almeida Viana (ISECENSA)

Rafaela Landim Gomes Siqueira (ISECENSA)

Pompilio Guimarães Reis Filho (ISECENSA)

Romeu e Silva Neto (ISECENSA)



Introdução: O desenvolvimento sustentável (DS) se tornou um conceito com várias concepções, devido principalmente à sua característica multidimensional, fato este que dificulta a sua operacionalização, e conseqüentemente, seu monitoramento e avaliação. **Objetivo:** Identificar padrões na composição dos resultados do Índice de Desenvolvimento Humano Municipal (IDHM), Índice Firjan de Desenvolvimento Municipal (IFDM) e Índice de Desenvolvimento Municipal Sustentável (IDMS) para os municípios brasileiros melhores ranqueados nessas três publicações. **Metodologia:** Selecionou-se uma amostra contendo os municípios dos estados do Rio de Janeiro e de Santa Catarina e aplicou-se a técnica de mineração de dados árvore de decisão às bases de dados correspondentes aos índices mencionados. **Resultados:** Indicam que os municípios catarinenses alcançaram posições mais expressivas nos três índices estudados e apresentaram menores variações de pontuação entre as dimensões desses índices. **Conclusões:** Sugere-se que composições de pontuação mais equilibradas entre as dimensões do desenvolvimento sejam capazes de promover melhores condições de desenvolvimento territorial.

Palavras-chave: Desenvolvimento sustentável, Mineração de dados, Árvore de decisão, Indicadores.

1. Introdução

“Um município é considerado mais ou menos sustentável à medida que é capaz de manter ou melhorar a saúde de seu sistema ambiental, minorar a degradação e o impacto antrópico, reduzir a desigualdade social e prover os habitantes de condições básicas de vida [...]” (BRAGA, 2004, p. 13). Esta multidimensionalidade torna o desenvolvimento sustentável (DS) um conceito com várias concepções (BELLEN, 2006). Sachs (1986) objetivou a discussão ao afirmar que o desenvolvimento sustentável (DS) só pode ser alcançado por meio do equilíbrio entre cinco dimensões, a saber: social, econômica, ecológica, geográfica e cultural, e Seiffert (2011) ressaltou que as dimensões do DS são extremamente inter-relacionadas e interdependentes e, portanto, devem ser analisadas de forma paralela e integrada pois ainda não se conhece o impacto dessas integrações (BELLEN 2006; VEIGA, 2010).

Porém, apesar do reconhecido papel que os indicadores de sustentabilidade dispõem frente às aspirações do DS, e considerando as várias iniciativas empreendidas com o intento de mensurar as suas dimensões, não é possível identificar razoável nível de convergência para que seja legitimado um indicador capaz de medir o DS de um território (VEIGA, 2009).

Como consequência da falta de consenso que circunda os cálculos da sustentabilidade, em suas dimensões, as formas atuais de medição do estágio de desenvolvimento utilizadas pelos municípios brasileiros ainda se concentram nos tradicionais Índice de Desenvolvimento Humano Municipal – IDHM (PNUD, 2013) e no Índice FIRJAN de Desenvolvimento Municipal – IFDM (FIRJAN, 2015). Além dos citados, o Índice de Desenvolvimento Municipal Sustentável – IDMS (FECAM, 2016), lançado em 2012, entrou no quadro dos indicadores sociais brasileiros e se apresentou como uma alternativa à mensuração do DS para os municípios do território nacional.

Devido à dificuldade em se estabelecer indicadores de desempenho apropriados aos domínios e objetivos definidos, surge a necessidade de se utilizar métodos e técnicas que melhorem o processo de seleção e monitoramento dos mesmos (PERAL, MATÉ e MARCO, 2017). Em contexto, a mineração de dados constitui uma importante ferramenta para auxiliar nesta tarefa, reduzindo a complexidade e o tempo gasto para examinar as relações entre os indicadores, seus resultados e suas características (SULTAN et al., 2017). Nesta direção, este trabalho possui como objetivo central identificar padrões na composição dos índices IDHM, IFDM e IDMS para os municípios brasileiros melhores ranqueados nessas três publicações, permitindo revelar a influência da integração das dimensões dos índices sobre as condições de desenvolvimento nestes municípios. Para tal, será aplicada a técnica de mineração de dados árvore de decisão às bases de dados correspondentes a estes três índices.

2. Mineração de Dados e Indicadores de Sustentabilidade

Com o objetivo de subsidiar a pesquisa com trabalhos recentes na literatura, foi realizada uma busca na base *Scopus* de trabalhos que abordassem os temas relacionados à pesquisa, envolvendo os termos “indicadores e índices”, “sustentabilidade” e “mineração de dados”. Isoladamente, os assuntos pesquisados retornaram um grande número de resultados, no entanto, quando combinados os três temas apenas 209 documentos foram retornados, o que demonstra certa escassez de trabalhos que unam os assuntos “indicadores e índices”, “sustentabilidade” e “mineração de dados” na literatura científica. Limitando os resultados a artigos e *conference papers*, e considerando apenas as publicações em língua inglesa, chegou-se a um quantitativo de 175 documentos, aos quais, inicialmente, tiveram seus títulos e resumos analisados a fim de se identificar alguma aderência dos trabalhos com a presente pesquisa. O intento da análise era identificar se esses trabalhos possuíam similaridades com a presente pesquisa, considerando os aspectos objetivo, objeto do estudo e metodologia. Ao final, foram então selecionados 14 documentos.

A mineração de texto como técnica de mineração de dados foi encontrada em seis dos catorze documentos resultantes. Modopothala e Isaac (2009) investigaram 2.415 Relatórios Ambientais Corporativos (RACs), por meio da mineração de texto, para medir a eficácia das RACs em termos de indicadores de desempenho econômico, ambiental e social, usando as diretrizes da *Global Reporting Initiative* (GRI). Rivera et al. (2014) define como desafiador a avaliação e o rastreamento dos indicadores de sustentabilidade, por se utilizarem de estudos caros e demorados. Nesta direção, os autores utilizaram a mineração de texto para explorar a viabilidade de identificação, rastreamento e descrição desses indicadores a partir de artigos não estruturados divulgados no meio digital. O trabalho de Milanez et al. (2014) teve como objetivo aplicar a mineração de texto a fim de verificar os avanços obtidos no ramo da nanocelulose a partir da investigação das patentes arquivadas no escritório de marcas e Patentes dos Estados Unidos entre os anos 2000 e 2012. Feil, de Quevedo e Schreiber (2015) utilização a mineração de texto para identificar e selecionar na literatura um conjunto de indicadores apropriados para medir a sustentabilidade industrial do setor de micro e pequenas fábricas de móveis. Os resultados apontaram para um grupo composto por 26 indicadores, sendo doze ambientais, sete sociais e sete econômicos.

Chanchetti et al. (2016) utilizaram a mineração de texto para analisar patentes na área de Materiais de Armazenamento de Hidrogênio (MACs), com o objetivo de avaliar o estágio do ciclo de vida dessas tecnologias, prever a tendência de desenvolvimento desses materiais e identificar os principais *stakeholders* envolvidos nesse contexto. Park e Kremer (2017)

identificaram que, apesar de existir uma grande quantidade de indicadores de sustentabilidade ambiental, os mesmos não eram organizados. Os autores então buscaram utilizar a mineração de texto para categorizar esses indicadores de acordo com formulários preenchidos por empresas no setor industrial e de serviços a respeito da utilização e utilidade de 55 indicadores de sustentabilidade ambiental encontrados na literatura.

Em relação aos trabalhos que aplicaram técnicas de mineração de dados em seus procedimentos técnicos, Barrera-Roldán et al. (2003) desenvolveu uma metodologia para construção de um índice de sustentabilidade industrial, a ser aplicado em uma refinaria na Cidade do Cabo. Definida a estrutura do indicador, os autores usaram a árvore de decisão para analisar os atributos gerais da árvore, que foram definidos como “avaliação financeira”, “econômico”, “social” e “ambiental”. Nijkamp, Rossi e Vindigni (2004) realizaram uma avaliação meta-analítica dos documentos encontrados na literatura que apresentavam valores estimados ou intervalos de valores do indicador *Ecological Footprint* (Pegada Ecológica). Foram empregadas técnicas de análise de frequência, tabulação cruzada e árvore de decisão nos documentos selecionados, e verificou-se que a escolha da técnica, dentre outros fatores, é um importante fator para explicar as variações nos resultados do indicador. Cheng e Wang (2008) utilizaram a mineração de dados históricos, em séries espaciais temporais, a fim de prevenir possíveis ocorrências de incêndios numa área florestal do Canadá.

Em Shareen et al. (2011) foi proposto um método para avaliação do grau de desenvolvimento de hidrocarbonetos em uma nação, dividido em cinco classes (futurista, conforme, sustentável, insustentável e crítica). O método foi capaz de rotular as classes não supervisionadas em cinco clusters distintos e a árvore de decisão foi utilizada para avaliar hierarquicamente o desenvolvimento do hidrocarboneto em um determinado país. Schönbrodt-Stitt et al. (2013) desenvolveram o modelo TerraCE (*Terrace Condition Erosion*), que utiliza a árvore de decisão, capaz de determinar as causas de diferentes condições de solo e seus processos de degradação através de pesquisas de campo e mineração de dados em séries espaciais. Buchert et al. (2015) salientou que durante o desenvolvimento de produtos sustentáveis, os engenheiros precisam prever as relações entre as características dos produtos e seus impactos econômicos, sociais e ambientais. Nesse sentido, foi proposto um modelo que integrasse o método de avaliação de sustentabilidade do ciclo de vida do produto e a árvore de decisão, de forma a sustentar o estudo de novos produtos e seus impactos nas dimensões citadas.

Zeng et al. (2016) desenvolveram um sistema de indicadores de sustentabilidade para cidades chinesas que possuíam atividades de mineração, onde foram coletados dados de 110 cidades espalhadas pelo território chinês. Destes dados, foram analisadas as características ambientais,

econômicas e sociais dessas cidades utilizando técnicas de associação de dados. Através da mineração dos dados, foi possível identificar características e padrões até então desconhecidos para as cidades mineradoras, como por exemplo o desequilíbrio entre os investimentos em educação, atividades econômicas e aspectos sociais nessas cidades. Já Hong, Jung e Park (2017) construíram um modelo de regressão preditiva, baseada em árvores de decisão, capaz de prever a dinâmica dos preços europeus das emissões de carbono com base nas movimentações históricas dos valores das *commodities* e dos produtos financeiros do mercado europeu.

3. Metodologia

3.1. Procedimentos técnicos

A metodologia utilizada neste artigo teve como objetivo sustentar a proposta de utilização da técnica de mineração de dados para identificar padrões na composição dos índices IDHM 2010, IFDM 2015 e IDMS 2016 dos municípios brasileiros. Nesta direção, primeiramente, foi realizada uma revisão da literatura nas bases *Scopus* (privada) e *Google Scholar* (pública) para exploração dos temas centrais desta pesquisa: os índices IDHM, IFDM e IDMS, e os demais índices que pudessem mensurar o desenvolvimento municipal, bem como técnicas de mineração de dados que possibilitassem identificar comportamentos e padrões em vetores de dados, com ênfase na técnica árvore de decisão. O passo seguinte caracterizou-se pela coleta geral de dados nas bases dos índices IDHM, IFDM e IDMS. Após, foi definida a amostra dos municípios a serem estudados no trabalho e então foram preparados os dados para aplicação da técnica. Na etapa seguinte aplicou-se a técnica de mineração de dados árvore de decisão nos dados preparados e, por fim, foram analisados os resultados obtidos.

3.2. Coleta de dados

A coleta dos índices de desenvolvimento IDHM 2010, IFDM 2015 e IDMS 2016 foi realizada por meio de consulta aos portais Atlas do Desenvolvimento Humano no Brasil (<http://www.atlasbrasil.org.br/2013/pt/consulta/>), Índice FIRJAN de Desenvolvimento Municipal (<http://www.firjan.com.br/ifdm/consulta-ao-indice/>) e ao Sistema de Indicadores de Desenvolvimento Municipal Sustentável (SIDEMS) (<http://indicadores.fecam.org.br/index/index/ano/2017>), respectivamente.

Na construção de cada índice é empregada uma metodologia específica, que considera suas próprias dimensões do desenvolvimento municipal e, portanto, fontes de dados distintas. No IDHM (PNUD, 2013) são utilizados os dados censitários do Instituto Brasileiro de Geografia e Estatística (IBGE), que compõem as dimensões “Vida longa e saudável”, “Acesso ao

conhecimento” e “Padrão de vida”, representando, respectivamente, os requisitos de longevidade, educação e renda. O IDHM varia de 0 a 1, possui classificação em cinco níveis (ou faixas), de muito baixo ao muito alto, respectivamente de 1 a 5. O IFDM (FIRJAN, 2015) é um indicador composto que utiliza os dados provenientes dos registros administrativos do Ministério do Trabalho e Emprego, do Ministério da Educação e do Ministério da Saúde, configurando as dimensões Emprego e Renda, Educação e Saúde do índice. O IFDM varia de 0 a 1, com quatro faixas de classificação, variando do nível baixo ao alto (1 a 4). O IDMS (FECAM, 2018) é formado por quatro dimensões, a saber Sociocultural, Econômica, Meio Ambiente e Político-institucional, e se utiliza de uma combinação de dados provenientes dos censos demográficos, das pesquisas de informações básicas municipais do IBGE e dos registros administrativos do Ministério da Saúde, do Ministério da Educação, do Ministério do Trabalho e Emprego, do Ministério da Fazenda, do Ministério das Cidades e do Tribunal Superior Eleitoral. A classificação do IDMS é dividida em cinco níveis, do baixo ao alto (1 a 5).

3.3. Definição da amostra e preparação dos dados

Conforme apresentado anteriormente, as publicações dos índices de desenvolvimento analisados possuem anos diferentes de lançamento e, por considerarem fontes distintas de dados, também possuem periodicidades de atualização diferentes, porém todos apresentam alta especificidade. A especificidade exprime o grau de consistência interna existente entre as dimensões e as variáveis constitutivas de um indicador, sendo uma propriedade particularmente importante para índices ou indicadores sociais compostos (JANUZZI, 2012).

O IDMS pode ser considerado como um índice com baixa factibilidade de obtenção, pois possui uma estrutura considerada complexa, composta por 4 dimensões, 9 subdimensões, 30 indicadores e 84 variáveis. Apesar de possuir ampla cobertura territorial, somente os resultados dos municípios localizados no Estado de Santa Catarina (ESC) são divulgados a nível das variáveis. Como forma de contornar esta restrição, este trabalho considerou como amostra exploratória o conjunto de municípios brasileiros localizados no ESC e no Estado do Rio de Janeiro (ERJ). Os resultados do IDMS para os municípios fluminenses foram calculados pelos próprios autores.

Definida a amostra, os dados coletados para os três índices foram organizados por faixas de classificação, de acordo com a metodologia de cada índice. Por exemplo, no IDHM 2010 o Município X obteve a nota 0,720, que segundo a classificação desse índice corresponde ao nível 4 (alto), logo foi considerado que o município atingiu o nível 4 de pontuação, e não a nota 0,720. Essa preparação dos dados foi necessária para a aplicação da técnica árvore de decisão.

3.4. Árvore de decisão

A árvore de decisão é uma das técnicas mais populares de mineração de dados, pois o seu uso como apoio à tomada de decisão e à descoberta de conhecimento gera resultados de fácil entendimento, em linguagem natural e semântica explícita, suportando objetivamente a justificativa de uma decisão. Esta técnica é comumente utilizada para resolução de problemas de atividades de classificação de dados, onde através de um algoritmo é realizada uma análise interativa dos atributos descritivos de um conjunto de dados rotulados, estabelecendo um processo de aprendizado do modelo classificador (SILVA, PERES, BOSCARIOLI; 2016). É fornecido um conjunto de objetos (vetor de atributos mais a classe) para o aprendizado do modelo, onde será analisada a relação entre os atributos e a classe dos objetos contidos no grupo de treinamento, para que então o modelo de classificação seja construído (GHEWARE, KEJKAR, TONDARE; 2014).

Uma árvore de decisão é uma representação hierárquica de relacionamentos organizados em uma estrutura semelhante a uma árvore convencional, cujo topo é ocupado por um nó raiz. Este nó raiz é então dividido em duas ou mais ramificações, de acordo com decisão tomada sobre o atributo avaliado, criando novos nós e ramos na árvore. Essa dinâmica ocorre até que alguma regra de parada seja satisfeita, como por exemplo um número mínimo de elementos por classe (NISBET, ELDER e MINER, 2009). Na literatura, muitos são os algoritmos que podem ser utilizados para aplicação desta técnica, sendo os tipos de atributos e as métricas de seleção desses atributos critérios que definem a escolha do algoritmo para indução da árvore de decisão (SILVA, PERES, BOSCARIOLI; 2016). Para o escopo deste trabalho foi escolhido o J48, baseado no algoritmo C4.5 (QUINLAN, 1993), “que lida com atributos do tipo categórico e numérico e, realiza um procedimento de poda que utiliza o raio do ganho de informação na seleção dos atributos que compõem os nós da árvore” (SILVA, PERES, BOSCARIOLI, 2016, p. 107). O software utilizado para a aplicação do J48 foi o *Waikato Environment for Knowledge Analysis* (WEKA), versão 3.8, disponível em: <https://www.cs.waikato.ac.nz/ml/weka/>.

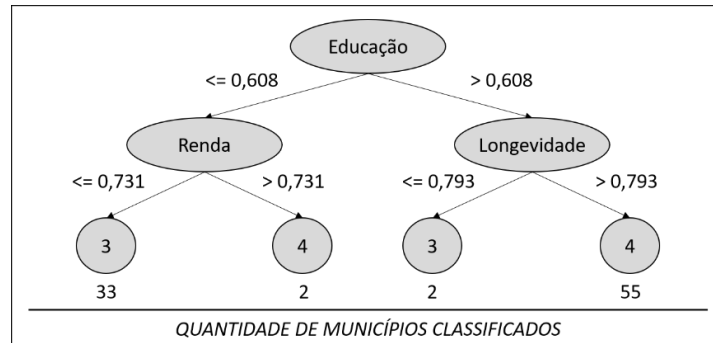
4. Resultados

4.1. IDHM 2010

Inicialmente foi aplicada a técnica da árvore de decisão para o conjunto dos municípios fluminenses e catarinenses segundo os resultados da publicação do IDHM 2010. Nota-se, conforme apresentado na Figura 1, que o atributo (dimensão) mais capaz de diferenciar o nível de classificação dos municípios do ERJ foi a dimensão Educação. Dos 92 municípios do ERJ classificados pela árvore de decisão, nenhum município classificado alcançou o nível mais alto

(5), 57 (62% do total) alcançaram o nível 4, sendo que 96% (55 municípios) dos que alcançaram o nível 4 pontuaram acima de 0,608 na dimensão Educação. E desses 96% dos municípios, todos tiveram pontuação acima de 0,793 na dimensão Longevidade. Mais de um terço dos municípios (36%) ocuparam as últimas posições do ranking, localizando-se na faixa 3 do IDHM fluminense.

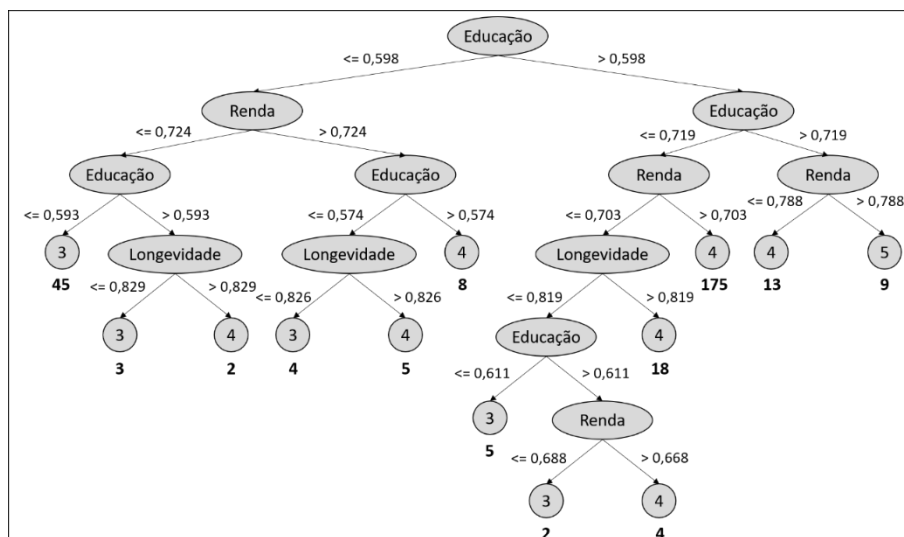
Figura 1 - Árvore de decisão para os municípios do ERJ no IDHM 2010.



Fonte: Os autores.

Para os municípios do ESC observou-se um aumento expressivo no número de nós da árvore gerada, fato este que já era esperado, visto o grande número de municípios localizados no território catarinense, 295, quando comparados aos 92 do ERJ. Dos 295 municípios, 293 foram classificados pela árvore de decisão e o atributo com maior capacidade de diferenciação do resultado da pontuação dos municípios também foi a dimensão Educação. Nove municípios foram classificados com nota cinco, o maior nível na classificação IDHM, todos com pontuação acima de 0,719 na Educação e acima de 0,788 na dimensão Renda. Apenas 20% (59 dos 293 classificados) dos municípios catarinenses obtiveram o IDHM no nível 3. A Figura 2 mostra a árvore de decisão resultante da classificação dos municípios do Estado de Santa Catarina no IDHM 2010.

Figura 2 - Árvore de decisão para os municípios do ESC no IDHM 2010.

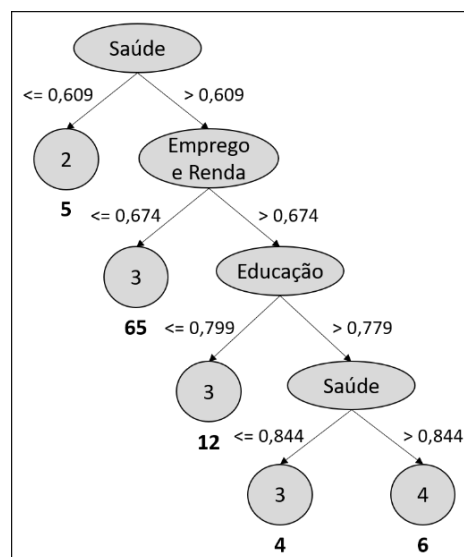


Fonte: Os autores.

4.2. IFDM 2015

Na segunda etapa de aplicação da mineração nos dados coletados, aplicou-se a técnica árvore de decisão ao conjunto dos municípios fluminenses e catarinenses nas bases do IFDM 2015. A árvore de decisão gerada através da aplicação da técnica aos resultados do IFDM 2015 para os municípios do ERJ está mostrada na Figura 3. Pode-se perceber que o atributo de maior capacidade de diferenciar o nível de resultado do índice foi a dimensão Saúde e apenas seis municípios fluminenses (7% dos 92 municípios do ERJ) alcançaram o maior nível (nota 4) no IFDM 2015. Todos esses obtiveram pontuação nessa dimensão acima de 0,844, na dimensão Emprego e Renda pontuação acima de 0,674 e na dimensão Educação resultados acima de 0,779. Apenas cinco municípios do ERJ (5% do total) tiveram IFDM classificados no nível 2.

Figura 3 - Árvore de decisão para os municípios do ERJ no IFDM 2015.

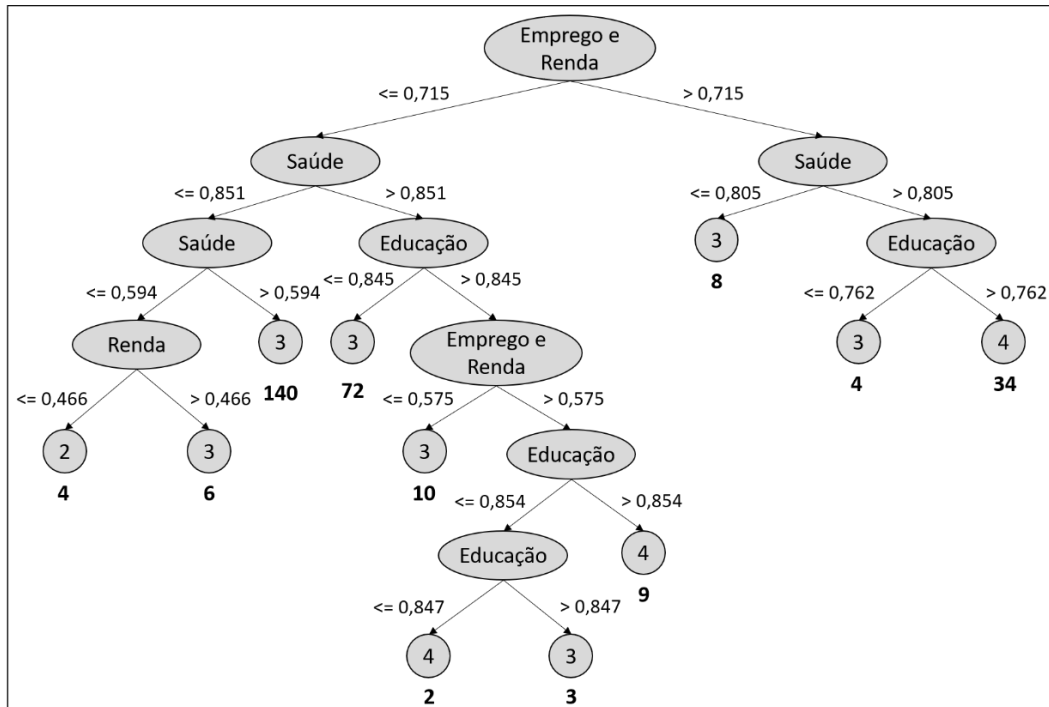


Fonte: Os autores.

Ao aplicar a técnica de mineração nos resultados do IFDM 2015 para os municípios do ESC também foi observado um aumento no número de nós da árvore de decisão gerada, ou seja, mais padrões de comportamento na composição do índice foram identificados. Duzentos e noventa e dois municípios catarinenses (99% do total) foram classificados pela técnica e a dimensão mais capaz de diferenciar as variáveis do modelo (as faixas de classificação do resultado do índice) foi a dimensão Emprego e Renda. Do total classificado, 45 municípios (15%) alcançaram a faixa mais alta do IFDM 2015 (nível 4) e, destes, 34 municípios tiveram pontuações maiores que 0,715 na dimensão Emprego e Renda, acima de 0,805 na Saúde e maiores que 0,762 na Educação. Nove dos 45 municípios com IFDM nível 4 obtiveram pontuações entre 0,575 e 0,715 na dimensão Emprego e Renda, acima de 0,851 na Saúde e acima de 0,854 na Educação. E dois dos 45 municípios com IFDM nível 4 apresentaram pontuações entre 0,575 e 0,715 na dimensão Emprego e Renda, acima de 0,851 na Saúde e entre 0,845 e 0,847 na dimensão Educação. Apenas 4 municípios do ESC (1% do total classificado)

ocuparam as últimas posições do ranking, localizando-se na faixa 2 do IFDM catarinense. A Figura 4 apresenta a árvore de decisão do IFDM 2015 para os municípios do ESC.

Figura 4 - Árvore de decisão para os municípios do ESC no IFDM 2015.

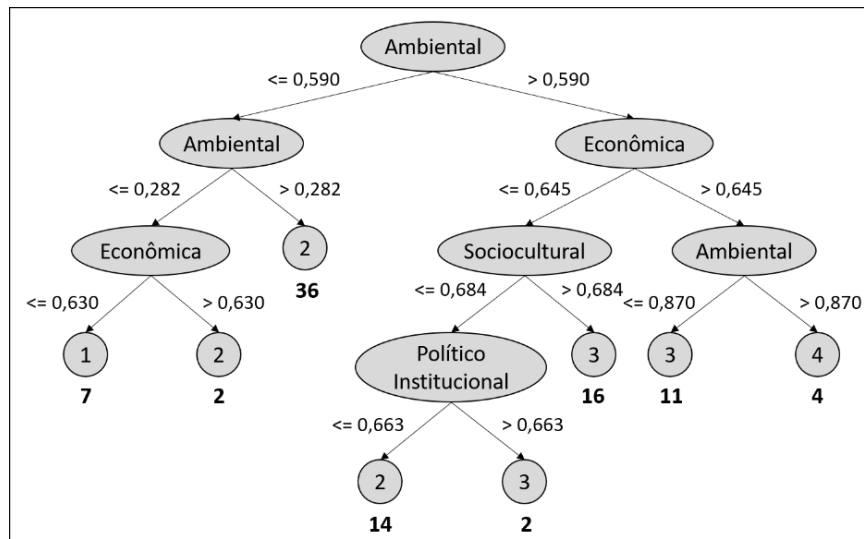


Fonte: Os autores.

4.3. IDMS 2016

Na terceira e última etapa de aplicação da mineração de dados nas bases coletadas, utilizou-se a técnica da árvore de decisão aos resultados do IDMS 2016 para os municípios do ERJ e do ESC. Como resultado, foram classificados os 92 municípios fluminenses na árvore de decisão e a dimensão que apresentou maior capacidade de diferenciação das faixas de pontuação do IDMS foi a Ambiental. Apenas quatro municípios (4% do total) se posicionaram no topo do ranking, porém ocupando o nível 4 do IDMS, um a menos que a melhor classificação do índice (nível 5). Esses 4 municípios alcançaram pontuações superiores a 0,870 na dimensão Ambiental e acima de 0,645 na dimensão Econômica. Sete municípios (8% do total) obtiveram a pior classificação do ranking e ocuparam o nível 1 no IDMS. A Figura 5 mostra a árvore de decisão gerada para os municípios fluminenses no IDMS 2016.

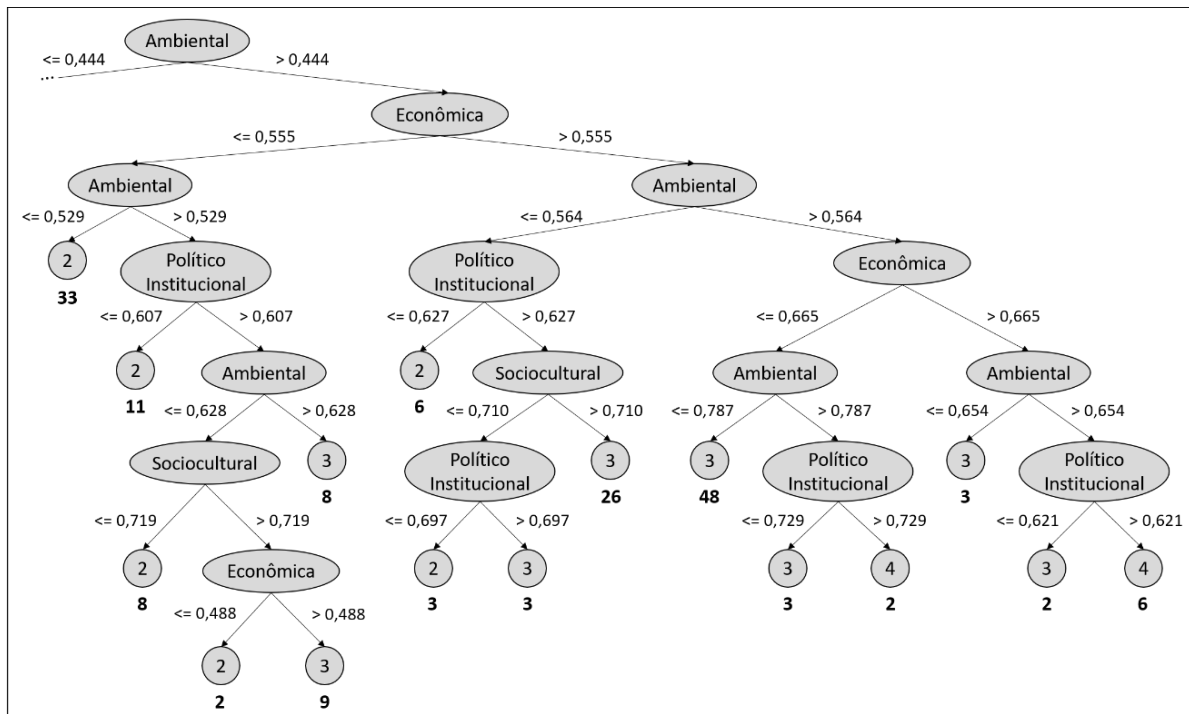
Figura 5 - Árvore de decisão para os municípios do ERJ no IDMS 2016.



Fonte: Os autores.

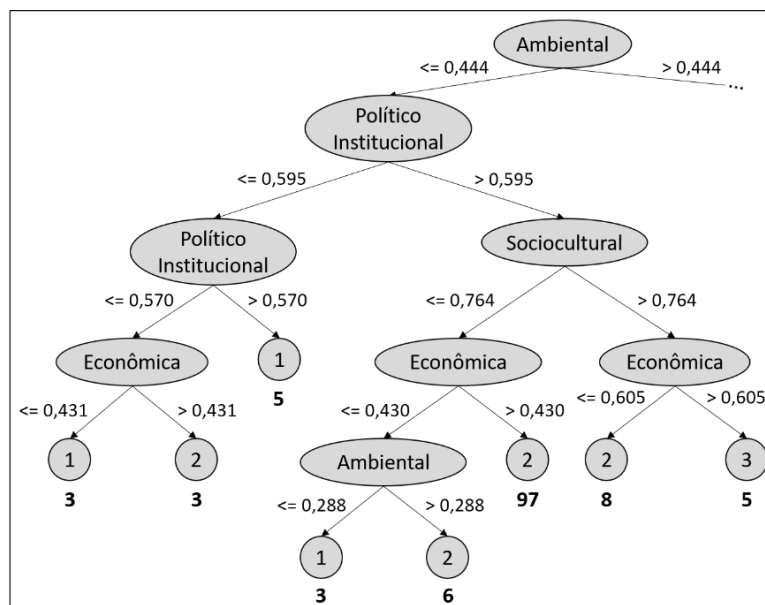
Por fim, foram classificados o conjunto dos 295 municípios catarinenses e o atributo mais significativo no que diz respeito à diferenciação das classes de pontuação novamente foi a dimensão Ambiental. Do total, nenhum município atingiu o nível máximo do IDMS e 8 municípios (3% do total) ocuparam as posições na faixa 4 de pontuação. Desses 8, 6 obtiveram pontuações superiores a 0,654 na dimensão Ambiental, acima de 0,665 na dimensão Econômica e maiores que 0,621 na dimensão Político-institucional. A Figura 6a exibe os ramos e nós da árvore de decisão, gerada a partir dos resultados do IDMS 2016 para os municípios de Santa Catarina, cuja pontuação na dimensão Ambiental é superior a 0,444. Na Figura 6b é apresentada a segunda parte da árvore de decisão dos municípios catarinenses, cujas pontuações no atributo Ambiental são iguais ou inferiores a 0,444. Nota-se que 8 municípios classificados figuraram entre as piores pontuações do IDMS 2016 e ocuparam o nível mais baixo de pontuação do índice (nível 1).

Figura 6a - Primeira parte da árvore de decisão para os municípios do ESC no IDMS 2016.



Fonte: Os autores.

Figura 6b - Segunda parte da árvore de decisão para os municípios do ESC no IDMS 2016.



Fonte: Os autores.

Um resumo dos resultados obtidos é mostrado no Quadro 1, composto pelas colunas de identificação: do índice (I); do Estado (II); do maior nível alcançado pelos municípios daquele Estado (III); da quantidade de municípios contida no melhor padrão detectado (IV); da porcentagem desses municípios em relação ao total de municípios classificados (V) e; dos intervalos de pontuação nas dimensões contidas no melhor padrão (VI). O asterisco na dimensão identifica o nó raiz daquela árvore.

Quadro 1 - Resumo dos resultados encontrados para os três índices.

| I | II | III | IV | V | VI | | | |
|--------------|----|-----|----|-----|---------------|-----------|-----------------|---------|
| | | | | | Longevidade | Educação | Renda | |
| IDHM 2010 | RJ | 4 | 55 | 60% | > 0,793 | > 0,608* | - | - |
| | SC | 5 | 9 | 3% | - | > 0,719* | > 0,788 | |
| | | | | | Saúde | Educação | Emprego e Renda | |
| IFDM 2015 | RJ | 4 | 6 | 7% | > 0,844* | > 0,779 | > 0,674 | |
| | SC | 4 | 34 | 12% | > 0,805 | > 0,762 | > 0,715* | |
| | | | | | Sociocultural | Econômica | Ambiental | |
| IDMS 2016 | RJ | 4 | 4 | 4% | - | > 0,645 | > 0,870* | - |
| | SC | 4 | 6 | 2% | - | > 0,665 | > 0,654* | > 0,621 |

Fonte: Os autores.

5. Conclusões e considerações finais

De forma geral, os resultados indicam que os municípios catarinenses alcançaram posições mais expressivas nos três índices estudados, quando comparados aos municípios fluminenses. No IDHM e no IDMS os municípios do ESC alcançaram o nível máximo pontuação, enquanto os municípios do ERJ não. E, apesar de posicionados na mesma faixa do IFDM, o ESC posicionou um número bem maior de municípios nesta faixa. Outro ponto a ser destacado reside no padrão dos resultados obtidos pelos municípios dos dois estados, havendo maior equilíbrio entre os intervalos de pontuação em todas as dimensões analisadas, por parte dos municípios catarinenses melhores classificados, com variações inferiores a um décimo, para os três índices. Este resultado sugere que composições mais uniformes de pontuação são capazes de atingir melhores classificações nos rankings das publicações, ou seja, maiores níveis de desenvolvimento. Para os gestores públicos, esses resultados aventam maior eficácia sobre formulações de políticas públicas que direcionem esforços, de maneira proporcional, sobre as dimensões do desenvolvimento. Nota-se também a pertinência da dimensão Ambiental, não presente no IDHM e IFDM, como o atributo de maior capacidade diferenciadora entre as classes de pontuação do IDMS.

Conforme apresentado anteriormente, poucos trabalhos exploraram propriamente os indicadores de sustentabilidade, e mesmo os que tiveram esta finalidade, optaram por empregar técnicas de mineração de texto em suas metodologias. Fato este que os diferencia integralmente do delineamento da presente pesquisa. Apesar de todos os trabalhos utilizarem técnicas de mineração de dados, apenas Nijkamp, Rossi e Vindigni (2004) e Zeng et al. (2016) tiveram como propósito investigar possíveis padrões encontrados na composição dos resultados dos

indicadores de sustentabilidade. No entanto, nenhum dos dois concentrou suas análises em índices sintéticos que mensurassem o desenvolvimento sustentável de um território. Tendo em face as razões apresentadas, torna-se inviável a discussão dos resultados encontrados pela presente pesquisa em relação às conclusões alcançadas pelos demais trabalhos pesquisados. Por este motivo, fica realçada a contribuição desta pesquisa ao utilizar a mineração de dados como forma de identificar padrões de comportamento nos resultados de um conjunto de índices de desenvolvimento sustentável para os municípios brasileiros.

Por fim, cabe destacar a relevância dos resultados encontrados, visto serem de grande valia para que sustentem os gestores municipais no que tange à formulação, implementação e avaliação de políticas públicas, capazes de promover um desenvolvimento de caráter sustentável e, por conseguinte equilibrado, em todas as suas dimensões. Sugere-se a realização de trabalhos futuros que contemplem demais municípios nacionais para que possam ser comparados os resultados com os obtidos nesta pesquisa, a fim de que seja enriquecida a discussão sobre possíveis caminhos que destinam à sustentabilidade de um território.

REFERÊNCIAS

- BARRERA-ROLDÁN, A. et al. Industrial sustainability index. **WIT Transactions on Ecology and the Environment**, v. 63, p. 337-346, 2003.
- BELLEN, H. M. V. **Indicadores de sustentabilidade: uma análise comparativa**. 2. ed. Rio de Janeiro: Editora FGV, 2006.
- BRAGA, T. M. et al. Índices de sustentabilidade municipal: o desafio de mensurar. **Nova Economia**, v. 14, n. 3, p. 11-33, 2004.
- BUCHERT, T. et al. Multi-criteria decision making as a tool for sustainable product development – Benefits and obstacles. **Procedia CIRP**, v. 26, p. 70-75, 2015.
- CHANCHETTI, L. F. et al. Technological forecasting of hydrogen storage materials using patent indicators. **International Journal of Hydrogen Energy**, v. 41, n. 41, p. 18301-18310, 2016.
- CHENG, T.; WANG, J. Integrated spatio-temporal data mining for forest fire prediction. **Transactions in GIS**, v. 12, n. 5, p. 591-611, 2008.
- FECAM. **Metodologia das variáveis do IDMS 2016**. Florianópolis, 2016.
- FEIL, A. A.; DE QUEVEDO, D. M.; SCHREIBER, D. Selection and identification of the indicators for quickly measuring sustainability in micro and small furniture industries. **Sustainable Production and Consumption**, v. 3, p. 34-44, 2015.
- FIRJAN. IFDM 2015 – Índice FIRJAN de Desenvolvimento Municipal. **Pesquisas e Estudos Socioeconômicos**, Rio de Janeiro, 2015.
- GHEWARE, S. D.; KEJKAR, A. S.; TONDARE, S. M. Data mining: task, tools, techniques and applications. **International Journal of Advanced Research in Computer and Communication Engineering**, v. 3, n. 10, p. 8095-8098, 2014.
- HONG, K.; JUNG, H.; PARK, M. Predicting european carbon emission price movements. **Carbon**

- Management**, v. 8, n. 1, p. 33-44, 2017.
- JANNUZZI, P. DE M. **Indicadores Sociais no Brasil**: conceitos, fontes de dados e aplicações. 5. ed. São Paulo: Editora Alínea, 2012.
- NIJKAMP, P.; ROSSI, E.; VINDIGNI, G. Ecological footprints in plural: a meta-analytic comparison of empirical results. **Regional Studies**, v. 38, n. 7, p. 747-765, 2004.
- NISBET, R.; ELDER, J.; MINER, G. **Handbook of statistical analysis and data mining applications**. San Diego: Elsevier, 2009.
- MILANEZ, D. H. et al. Technological indicators of nanocellulose advances obtained from data and text mining applied to patent documents. **Materials Research**, v. 17, n. 6, p. 1513-1522, 2014.
- MODAPOTHALA, J. R.; ISSAC, B. Study of economic, environmental and social factors in sustainability reports using text mining and Bayesian analysis. In: **IEEE SYMPOSIUM ON INDUSTRIAL ELECTRONICS & APPLICATIONS (ISIEA)**, 2009, Kuala Lumpur. p. 209-214.
- PARK, K.; KREMER, G. E. O. Text mining-based categorization and user perspective analysis of environmental sustainability indicators for manufacturing and service systems. **Ecological Indicators**, v. 72, p. 803-820, 2017.
- PERAL, J.; MATÉ, A.; MARCO. Application of data mining techniques to identify relevant key performance indicators. **Computer Standards & Interfaces**, v. 50, parte 2, p. 55-64, 2017.
- PNUD. Índice de Desenvolvimento Humano Municipal Brasileiro. **Atlas do Desenvolvimento Humano no Brasil**. Brasília, Distrito Federal, 2013.
- QUINLAN, J. R. *C4.5: Programs for Machine Learning*. San Francisco: Morgan Kaufmann Publishers Inc., 1993.
- RIVERA, S. J. et al. A text mining framework for advancing sustainability indicators. **Environmental modelling & software**, v. 62, p. 128-138, 2014.
- SACHS, I. **Ecodesenvolvimento**: crescer sem destruir. São Paulo: Vértice, 1986.
- SCHÖNBRODT-STITT, S. et al. Degradation of cultivated bench terraces in the Three Gorges Area: field mapping and data mining. **Ecological indicators**, v. 34, p. 478-493, 2013.
- SEIFFERT, M. E. B. **Gestão Ambiental**: instrumentos, esferas de ação e educação ambiental. 2. ed. São Paulo: Atlas, 2011.
- SHAHEEN, M. et al. Mining sustainability indicators to classify hydrocarbon development. **Knowledge-Based Systems**, v. 24, n. 8, p. 1159-1168, 2011.
- SILVA, L. A. DA; PERES, S. M.; BOSCARIOLI, C. **Introdução à mineração de dados**: com aplicações em R. Rio de Janeiro: Elsevier, 2016.
- SULTAN, N. et al. Data mining approach for detecting key performance indicators. **Journal of Artificial Intelligence**, v. 10, n. 2, p. 59-65. 2017.
- VEIGA, J. E. DA. Indicadores socioambientais: evolução e perspectivas. **Revista de Economia Política**, v. 29, n. 4, p. 421-435, 2009.
- VEIGA, J. E. DA. **Desenvolvimento sustentável**: o desafio do século XXI. Rio de Janeiro: Garamond, 2010.
- ZENG, L. et al. Analyzing sustainability of chinese mining cities using an association rule mining approach. **Resources Policy**, v. 49, p. 394-404, 2016.