

**Instituto Federal de Educação, Ciência e
Tecnologia Fluminense**
**Programa de Pós-graduação em Sistemas Aplicados à
Engenharia e Gestão**

**ATRIBUTOS DISCENTES RELEVANTES NO RESULTADO DO PISA:
MINERAÇÃO DE DADOS APLICADA NA AVALIAÇÃO EDUCACIONAL**

ROBERTA ALVARENGA DOS SANTOS

2018

Instituto Federal de Educação, Ciência e Tecnologia Fluminense
Programa de Pós-graduação em Sistemas Aplicados à Engenharia e Gestão

**ATRIBUTOS DISCENTES RELEVANTES NO RESULTADO DO PISA:
MINERAÇÃO DE DADOS APLICADA NA AVALIAÇÃO EDUCACIONAL**

ROBERTA ALVARENGA DOS SANTOS

Henrique Rego Monteiro da Hora
(Orientador)

Dissertação submetida como requisito parcial para obtenção do grau de **Mestre** no Programa de Pós-graduação em Sistemas Aplicados à Engenharia e Gestão, Área de Concentração em Sistemas Computacionais.

Campos dos Goytacazes, RJ
Dezembro de 2018

Biblioteca Anton Dakitsch
CIP - Catalogação na Publicação

S237a

Santos, Roberta Alvarenga dos
ATRIBUTOS DISCENTES RELEVANTES NO RESULTADO DO
PISA: MINERAÇÃO DE DADOS APLICADA NA AVALIAÇÃO
EDUCACIONAL / Roberta Alvarenga dos Santos - 2018.
66 f.: il. color.

Orientador: Henrique Rego Monteiro da Hora

Dissertação (mestrado) -- Instituto Federal de Educação, Ciência e
Tecnologia Fluminense, Campus Campos Centro, Curso de Mestrado
Profissional em Sistemas Aplicados à Engenharia e Gestão, Campos dos
Goytacazes, RJ, 2018.

Referências: f. 65 a 66.

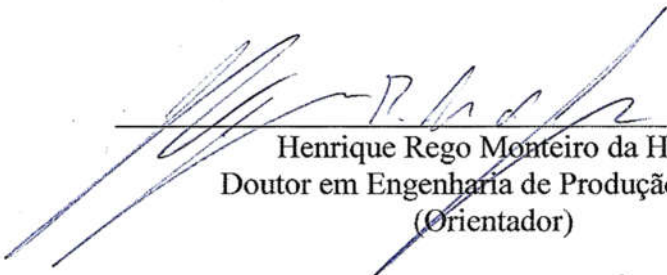
1. Avaliação. 2. Educação. 3. PISA. 4. Indicadores educacionais. I. Hora,
Henrique Rego Monteiro da , orient. II. Título.

**Instituto Federal de Educação, Ciência e
Tecnologia Fluminense**
Programa de Pós-graduação em Sistemas Aplicados à Engenharia e Gestão

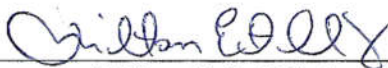
ROBERTA ALVARENGA DOS SANTOS

Dissertação submetida como requisito parcial para obtenção do grau de **Mestre** no Programa de Pós-graduação em Sistemas Aplicados à Engenharia e Gestão, Área de Concentração em Sistemas Computacionais.

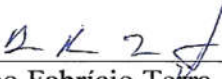
APRESENTADO EM 21/12/2018



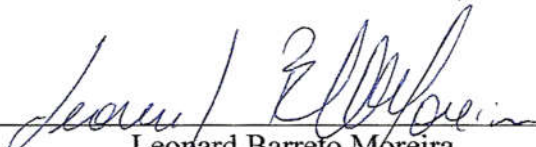
Henrique Rego Monteiro da Hora
Doutor em Engenharia de Produção - UFF
(Orientador)



Milton Erthal Júnior
Doutor em Produção Vegetal – UENF



Breno Fabrício Terra Azevedo
Doutor em Informática na Educação – UFRGS



Leonard Barreto Moreira
Doutor em Modelagem Computacional - IPRJ

AGRADECIMENTOS

A Deus, pela dádiva da vida, por me guiar, iluminar e me dar tranquilidade para seguir em frente com os meus objetivos e não desanimar com as dificuldades.

A minha família, em particular, meus irmãos Renata e Rodrigo, meus sobrinhos Arthur, Alice e Ana, meus cunhados Gabriela e Gustavo, mas de modo muito especial a minha mãe, Valdinéa, por sempre acreditar em mim e por ter abdicado de sua vida em prol das realizações e da felicidade de seus filhos.

Ao meu pai, Renato (*in memoriam*), que em algum lugar deve estar feliz com a minha vitória.

Ao Cássio, pelo seu amor, carinho, auxílio, paciência, pelas palavras sábias e pela presença incansável com que me apoiou ao longo de todo o período do mestrado. Sem você do meu lado esse trabalho não seria possível. Agradeço também a sua família, pela amizade e apoio.

Ao Prof. Henrique da Hora, pela orientação, incentivo, disponibilidade e dedicação tão importantes. Obrigado por acreditar e exigir de mim muito mais do que eu imaginava ser capaz de fazer.

Aos membros da banca examinadora, Prof. Milton, Prof. Breno e Prof. Leonard, que tão gentilmente aceitaram participar e colaborar com esta dissertação.

Ao aluno bolsista Rosemilson, pelo auxílio na obtenção e tratamento da base de dados.

Aos professores e alunos da turma I e II do SAEG, pela amizade e companheirismo.

Aos meus amigos, pela compreensão do tempo de convívio, muitas vezes sacrificado para realização deste trabalho.

A todos os meus colegas e alunos do CENSA, Externato Campista, C.R. Dr. Thiers Cardoso, UENF, ISEPAM e EEEFM Presidente Kennedy, principalmente a equipe de coordenação e gestão, pelo incentivo e por serem tão solícitos às minhas necessidades para que pudesse desempenhar paralelamente ao mestrado o meu papel de docente.

O meu profundo e sincero agradecimento a todas as pessoas que contribuíram para a concretização desta dissertação, estimulando-me intelectual e emocionalmente.

LISTA DE FIGURAS

Figura 3.1 – Visão geral das etapas que compõe o processo KDD	11
Figura 3.2 – Diagrama de <i>Venn</i> com a quantidade de trabalhos encontrados na base <i>Scopus</i>	12
Figura 3.3 – Evolução cronológica.	13
Figura 3.4 – Autores mais citados.....	13
Figura 3.5 – Periódicos e conferências nas áreas ABC, BCD e ACD.	15
Figura 3.6 – Periódicos e conferências na área ABD.	16
Figura 3.7 – Diagrama de <i>Venn</i> atualizado com a quantidade de trabalhos encontrados na base <i>Scopus</i>	19
Figura 4.1 – Visão geral das etapas que compõe o processo KDD.	24
Figura 4.2 – Etapas da metodologia da pesquisa.	27
Figura 4.3 – Mapa de árvore dos temas relevantes.	46
Figura 4.4 – Árvore de decisão de Singapura: Matemática.	57
Figura 4.5 – Árvore de decisão de Hong Kong: Matemática.	57
Figura 4.6 – Árvore de decisão de Macau (China): Matemática.	58
Figura 4.7 – Árvore de decisão de Taipei Chinesa: Matemática.	58
Figura 4.8 – Árvore de decisão de Japão: Matemática.	59
Figura 4.9 – Árvore de decisão de Singapura: Leitura.	59
Figura 4.10 – Árvore de decisão de Hong Kong: Leitura.....	60
Figura 4.11 – Árvore de decisão de Canadá: Leitura.....	60
Figura 4.12 – Árvore de decisão de Finlândia: Leitura.	61
Figura 4.13 – Árvore de decisão de Irlanda: Leitura.	61
Figura 4.14 – Árvore de decisão de Singapura: Ciências.	62
Figura 4.15 – Árvore de decisão de Japão: Ciências.	62
Figura 4.16 – Árvore de decisão de Estônia: Ciências.	63

Figura 4.17 – Árvore de decisão de Taipei Chinesa: Ciências.	63
Figura 4.18 – Árvore de decisão de Finlândia: Ciências.	64

LISTA DE QUADROS

Quadro 2.1 – Instrumentos utilizados pelo PISA	3
Quadro 2.2 – Países e economias do PISA 2015.....	4
Quadro 3.1 – Termos utilizados para a pesquisa.	12
Quadro 3.2 – Trabalhos mais citados.	14
Quadro 4.1 – Atributos gerados nas árvores de decisão: Matemática.	30
Quadro 4.2 – Descrição dos Atributos gerados nas árvores de decisão: Matemática.	31
Quadro 4.3 – Atributos gerados nas árvores de decisão: Leitura.	36
Quadro 4.4 – Descrição dos Atributos gerados nas árvores de decisão: Leitura.....	37
Quadro 4.5 – Atributos gerados nas árvores de decisão: Ciências.	41
Quadro 4.6 – Descrição dos Atributos gerados nas árvores de decisão: Ciências.	41

LISTA DE TABELAS

Tabela 2.1 – Notas do PISA 2015.....	5
Tabela 4.1 – Amostra utilizada na pesquisa.....	23
Tabela 4.2 – Taxas de acerto para os países com alto desempenho em Matemática.....	30
Tabela 4.3 – Taxas de acerto para os países com alto desempenho em Leitura	36
Tabela 4.4 – Taxas de acerto para os países com alto desempenho em Ciências	40

SUMÁRIO

1. APRESENTAÇÃO	1
2. O PROGRAMA INTERNACIONAL DE AVALIAÇÃO DE ESTUDANTES (PISA)	3
3. ARTIGO I: ESTUDO BIBLIOMÉTRICO SOBRE AVALIAÇÃO EDUCACIONAL UTILIZANDO MINERAÇÃO DE DADOS NA BASE DO TESTE DE PISA.....	6
3.1. Resumo	6
3.2. Abstract.....	6
3.3. Introdução	7
3.4. Metodologia.....	8
3.5. Referencial Teórico	9
3.5.1. Estudo Bibliométrico	9
3.5.2. Scopus.....	10
3.5.3. Avaliação Educacional e o PISA	10
3.5.4. Mineração de Dados (DM).....	11
3.6. Resultados.....	11
3.7. Conclusão	16
Referências Bibliográficas	17
Apêndice A	19
4. ARTIGO II: IDENTIFICAÇÃO DOS ATRIBUTOS DISCENTES RELEVANTES NO RESULTADO DO PISA UTILIZANDO TÉCNICAS DE MINERAÇÃO DE DADOS.....	20
4.1. Resumo	20
4.2. Abstract.....	20
4.3. Introdução	21
4.4. Metodologia.....	23
4.4.1. Classificação da pesquisa	23
4.4.2. População e amostra	23

4.4.3. Pesquisa metodológica	24
4.4.4. Procedimentos técnicos	27
4.5. Resultados.....	29
4.5.1. Países com alto desempenho em Matemática.....	30
4.5.2. Países com alto desempenho em Leitura.....	36
4.5.3. Países com alto desempenho em Ciências.....	40
4.6. Análise e Discussão de Resultados.....	46
4.7. Conclusão	52
Referências Bibliográficas	53
Apêndice A	57
5. CONSIDERAÇÕES FINAIS.....	65
REFERÊNCIAS BIBLIOGRÁFICAS.....	65

1. APRESENTAÇÃO

A produtividade econômica de um país está intimamente ligada ao seu sistema educacional e a melhora do nível de educação de uma nação promove a formação de trabalhadores mais flexíveis na sua capacitação profissional (BARONE, 2017). Assim, revela-se a importância de se discutir as políticas a serem adotadas para criar ou otimizar mecanismos que fomentam a educação numa nação. Essa necessidade é mais evidente em países que são conhecidos pela baixa qualidade historicamente identificada em seus sistemas de ensino.

Nesse sentido, as avaliações educacionais exercem o papel de auxiliar a compreensão do funcionamento dos sistemas educacionais de um país, buscando identificar os pontos positivos e negativos dos mecanismos vigentes. Assim, em 1997, foi criado o Programa Internacional de Avaliação de Estudantes (PISA), da Organização para a Cooperação e Desenvolvimento Econômico (OCDE). Ele é aplicado para estudantes espalhados por mais de 70 países no mundo. Seu principal objetivo é gerar indicadores que permitam definir políticas de grande escala a partir das forças e fraquezas dos sistemas educacionais dos diferentes países (MARTÍNEZ; CHARRO, 2018). Um exemplo de aplicação da análise dos dados gerados no teste PISA pode ser visto em Tastan e Erdoğan (2018) e Teltemann e Windzio (2018).

Na avaliação do PISA, os alunos buscam resolver uma sequência de questões com nível crescente de dificuldade. Além disso, cada estudante, pais e responsáveis pela escola preenchem um questionário básico. Esses dados são coletados e reunidos num conjunto de dados rico em informações. Assim, técnicas de extração de conhecimento têm sido cada vez mais estudadas e aplicadas na base de dados gerada pela aplicação do PISA (PEJIC; MOLCER, 2018).

Dentre essas técnicas, destaca-se a mineração de dados (*Data Mining* – DM), cujo estudo e aplicação tem sido crescente. Ela consiste no processamento de grandes quantidades de dados, transformando-as em conhecimento útil, de forma inteligente e automática (REIS, 2015). Os principais métodos de DM mais utilizados para examinar bancos de dados educacionais são os de classificação, clusterização, análises de regressão e regras de associação (ROMERO, 2011).

Segundo a OCDE (2016), os dados do PISA ajudam a monitorar as tendências na aquisição de conhecimentos e habilidades por parte dos estudantes nos países e em diferentes subgrupos demográficos dentro de cada país. As descobertas permitem que os formuladores de políticas em todo o mundo avaliem os conhecimentos e as habilidades dos estudantes em seus próprios países, em comparação com os de outros países; estabeleçam metas políticas contra objetivos mensuráveis

alcançados por outros sistemas educacionais; e aprendam com políticas e práticas aplicadas em outros lugares.

Peri *et al.* (2016) afirmaram que, na educação, a comparação dos resultados é um objetivo ambicioso e representa muitos desafios. A possibilidade de gerar comparações rígidas entre os sistemas educacionais constitui um dos principais pontos do PISA, pois o teste fornece informações válidas e confiáveis não só sobre os fatores que influenciam no desempenho dos alunos, mas também sobre as características estruturais dos sistemas educacionais.

Com base neste contexto, a questão investigada no presente trabalho é:

Quais são os fatores que foram relevantes para o sucesso educacional dos países bem-sucedidos na avaliação do PISA?

O objetivo deste trabalho é extrair conhecimento da base dados do PISA, utilizando técnicas de DM, de forma a se analisar quais são os fatores que foram relevantes para o sucesso educacional dos países que se destacaram na avaliação. Foram definidos os seguintes objetivos específicos: Mapear a produção científica na área de avaliação educacional com foco no teste do PISA que utilizam técnicas de mineração de dados; aplicar técnicas de mineração de dados na base de dados do PISA para identificar padrões em potencial que expliquem o sucesso dos países que se destacaram na avaliação.

A presente pesquisa está estruturada em cinco capítulos. Este primeiro capítulo traz uma apresentação do tema, juntamente com a definição da questão de pesquisa. No segundo capítulo é apresentado o Programa Internacional de Avaliação de Estudantes (PISA) e seus reflexos no Brasil. No terceiro capítulo é apresentado um estudo bibliométrico sobre a produção científica na área de avaliação educacional com foco no teste do PISA que utilizam técnicas de DM. No quarto capítulo, é realizado a identificação e discussão dos atributos que foram relevantes para o sucesso educacional dos países que se destacaram na avaliação do PISA. Por fim, o quinto capítulo traz as considerações finais deste trabalho, seguido pelas referências bibliográficas utilizadas na pesquisa.

2. O PROGRAMA INTERNACIONAL DE AVALIAÇÃO DE ESTUDANTES (PISA)

Na busca de responder à pergunta "O que é importante para os cidadãos conhecerem e serem capazes de fazer?", e também atendendo a necessidade de criar resultados internacionalmente comparáveis sobre o desempenho dos alunos, a OCDE lançou a pesquisa trienal de estudantes de 15 anos em todo o mundo, conhecido como PISA (OCDE, 2016).

A primeira aplicação do PISA foi em 2000 e a definição de ser aplicado a estudantes com idade de 15 anos foi decidida em comum acordo entre a OCDE e os países participantes, levando em consideração de que essa é idade média em que os jovens, na maioria dos países da OCDE, estão próximos do término da educação compulsória (OCDE, 2015). Segundo OCDE (2016), a avaliação centra-se nas principais matérias da escola de Matemática, Leitura e Ciências e não apenas determina se os alunos podem reproduzir o conhecimento, mas também examina o quão bem os alunos podem extrapolar do que aprenderam e podem aplicar esse conhecimento, tanto dentro como fora da escola.

Em cada edição é selecionada uma das áreas cognitivas como foco principal da avaliação, sendo a maior parte dos itens centrada nessa área e os demais voltados para as outras duas áreas. Os instrumentos do PISA fornecem três principais tipos de resultados: indicadores, que fornecem um perfil básico de conhecimento e habilidades dos estudantes; indicadores derivados de questionários, que mostram como tais habilidades são relacionadas a variáveis sociais, econômicas, demográficas e educacionais; e indicadores de tendências, que acompanham o desempenho dos estudantes e monitoram os sistemas educacionais ao longo do tempo (INEP, 2016). O Quadro 2.1 apresenta os instrumentos utilizados pelo PISA desde sua primeira edição, ressaltando a área cognitiva, que foi o foco principal em cada ano de aplicação.

Quadro 2.1 – Instrumentos utilizados pelo PISA

PISA	2000	2003	2006	2009	2012	2015
Testes	Leitura	Leitura	Leitura	Leitura	Leitura	Leitura
	Matemática	Matemática	Matemática	Matemática	Matemática	Matemática
	Ciências	Ciências	Ciências	Ciências	Ciências	Ciências
						Resolução colaborativa de problemas
						Letramento financeiro
Questionários	Estudante – Geral	Estudante – Geral	Estudante – Geral	Estudante – Geral	Estudante – Geral	Estudante – Geral
	Escola	Escola	Escola	Escola	Escola	Escola
						Estudante – Familiaridade com tecnologia
						Professor

Fonte: INEP (2016).

Em 2015, o foco principal do PISA foi em Ciências, ficando a Leitura, Matemática, como áreas menores de avaliação. O PISA 2015 também incluiu mais duas avaliações, uma de resolução colaborativa de problemas e outra, opcional, de alfabetização financeira. Cerca de 540000 alunos completaram a avaliação, representando cerca de 29 milhões de jovens de 15 anos nas escolas dos 72 países e economias participantes (OCDE, 2016). Destes, 35 são países membros da OCDE e 37 são países/economias parceiros em que a avaliação foi aplicada, conforme mostrado no Quadro 2.2.

Quadro 2.2 – Países e economias do PISA 2015

Países OCDE	Países e economias parceiras PISA 2015
Alemanha, Austrália, Áustria, Bélgica, Canadá, Chile, Coreia do Sul, Dinamarca, Eslováquia, Eslovênia, Espanha, Estados Unidos, Estônia, Finlândia, França, Grécia, Holanda, Hungria, Irlanda, Islândia, Israel, Itália, Japão, Letônia, Luxemburgo, México, Noruega, Nova Zelândia, Polônia, Portugal, Reino Unido, República Tcheca, Suécia, Suíça e Turquia.	Brasil, Albânia, Argélia, Argentina, Bulgária, Catar, Cazaquistão, Cingapura, Colômbia, Costa Rica, Croácia, Emirados Árabes Unidos, Geórgia, Hong Kong, Indonésia, Jordânia, Kosovo, Líbano, Lituânia, Macau, Macedônia, Malásia, Malta, Moldávia, Montenegro, Peru, República Dominicana, Romênia, Rússia, Tailândia, Taipei, Trinidad e Tobago, Tunísia, Uruguai e Vietnã.

Fonte: INEP (2016).

No Brasil, o PISA faz parte de um conjunto de avaliações e exames nacionais e internacionais coordenados pela Diretoria de Avaliação da Educação Básica (DAEB), do Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (INEP) (INEP, 2016). O relatório do PISA 2015 mostrou que o desempenho dos alunos no Brasil está abaixo da média dos alunos em países da OCDE em Ciências (401 pontos, comparados à média de 493 pontos), em Leitura (407 pontos, comparados à média de 493 pontos) e em Matemática (377 pontos, comparados à média de 490 pontos), conforme mostra a Tabela 2.1 (OCDE, 2016).

Tabela 2.1 – Notas do PISA 2015.

Ciências 2015			Leitura 2015			Matemática 2015		
1	Singapura	556	1	Singapura	535	1	Singapura	564
2	Japão	538	2	Hong Kong	527	2	Hong Kong	548
3	Estônia	534	3	Canadá	527	3	Macau (China)	544
4	Taipei Chinesa	532	4	Finlândia	526	4	Taipei Chinesa	542
5	Finlândia	531	5	Irlanda	521	5	Japão	532
Média OCDE		493	Média OCDE		493	Média OCDE		490
59	Montenegro	411	59	Brasil	407	59	Costa Rica	400
60	Geórgia	411	60	Albânia	405	60	Líbano	396
61	Jordânia	409	61	Catar	402	61	Colômbia	390
62	Indonésia	403	62	Geórgia	401	62	Peru	387
63	Brasil	401	63	Peru	398	63	Indonésia	386
64	Peru	397	64	Indonésia	397	64	Jordânia	380
65	Líbano	386	65	Tunísia	361	65	Brasil	377
66	Tunísia	386	66	Rep. Dom.	358	66	ARIM	371
67	ARIM	384	67	ARIM	352	67	Tunísia	367
68	Kosovo	378	68	Argélia	350	68	Kosovo	362
69	Argélia	376	69	Kosovo	347	69	Argélia	360
70	Rep. Dom.	332	70	Líbano	347	70	Rep. Dom.	328

Fonte: Elaboração própria (2018).

Muri e Ortigão (2017), afirmaram que o Brasil é colocado no nível 1 da escala de desempenho PISA, mas ainda existe uma margem significativa de estudantes brasileiros que não conseguem alcançar isso, que é o nível mais baixo considerado pelo programa. Os resultados obtidos pelos estudantes brasileiros no PISA colocaram o país em uma posição desvantajosa comparada a praticamente todos os países participantes do teste. Os autores ressaltaram que o desenvolvimento científico e tecnológico do Brasil exige uma grande revisão da estrutura educacional adotada. Complementando, Ferrer (2017), ressaltou que os dados fornecidos pela avaliação do PISA permitem aos gestores políticos observar quais fatores estão associados ao sucesso educacional de alguns países, estabelecendo comparações entre resultados e estruturando seus sistemas afim de melhorar a qualidade educacional e alcançar resultados positivos.

Sendo assim, o presente trabalho justifica-se por explorar a área de mineração de dados educacionais, com a finalidade de contribuir para o debate em avaliação educacional, fomentando novas pesquisas e, principalmente, influenciando a promoção do uso dos resultados do PISA no Brasil, na busca de melhorias para a educação do país.

3. ARTIGO I: ESTUDO BIBLIOMÉTRICO SOBRE AVALIAÇÃO EDUCACIONAL UTILIZANDO MINERAÇÃO DE DADOS NA BASE DO TESTE DE PISA

3.1. Resumo

A avaliação educacional constitui um elemento chave para melhorar a qualidade de qualquer escola ou, em seu sentido mais amplo, a qualidade do sistema educacional como um todo. Uma avaliação educacional é importante não só para avaliar a qualidade da educação oferecida, mas também para identificar os pontos fortes e fracos, procurando garantir que todas as escolas ofereçam aos alunos uma educação de qualidade que atenda aos padrões de excelência alcançados nos países mais bem desenvolvidos na questão educacional. O Programa Internacional de Avaliação de Estudantes (PISA) é um programa internacional de testes da Organização para a Cooperação e Desenvolvimento Econômico (OCDE) realizado em mais de 60 países em todo o mundo. O banco de dados é de acesso aberto, e, portanto, oferece uma possibilidade para exploração científica e mineração de dados. A exigência de um estudo aprofundado dos dados educacionais pressupõe o emprego de tecnologias que possibilitem analisar uma vasta quantidade de dados e, por conseguinte, extrair informações e conhecimento relevante. A mineração de dados se apresenta como uma das técnicas que auxilia a descoberta de conhecimento em grandes bases de dados. Um estudo bibliométrico nessa área é relevante por fornecer informações e orientações para pesquisadores em relação ao andamento da produção científica. O objetivo dessa pesquisa é mapear a produção científica na área de avaliação educacional com foco no teste do PISA, utilizando técnicas de mineração de dados, com o objetivo de apresentar uma visão mais ampla da produção científica nacional e internacional dentro deste contexto. Para tanto, foi realizado um estudo bibliométrico da produção científica indexada na base de dados *Scopus*, considerando o período dos últimos dez anos. Foi possível verificar que a área de avaliação educacional usando mineração de dados se destacou em termos de quantidade de publicações, apresentando uma produção crescente ao longo do período avaliado. Os resultados favoreceram ao conhecimento de um cenário mais abrangente sobre a produção científica sobre o tema e apontaram para a ausência de trabalhos que empregam a mineração de dados na base do PISA para realização de uma avaliação educacional. Isso indica um *déficit* de produção nessa área, que tem grande potencial a ser explorada.

Palavras-chave: Indicadores educacionais. Programa Internacional de Avaliação de Estudantes. Mineração de dados.

3.2. Abstract

The educational evaluation constitutes a key element to improve the quality of any school or, in its broadest sense, the quality of the education system as a whole. The process of conducting educational evaluation is important not only to evaluate the quality of schools, but also to identify their strengths and weaknesses, ensure that all schools offer students a quality education that meets standards of excellence achieved in the most well-developed countries on the educational issue. The Programme for International Student Assessment (PISA) is an international testing program of Organisation for Economic Co-operation and Development (OECD) held in over 60 countries all over the world. The database is open access, and hence it offers a possibility for scientific exploration and data mining. The requirements of a detailed study of educational data presuppose the use of technologies that allow the analysis of a vast amount of data, and therefore, extract relevant information and

knowledge. The data mining is one of the techniques that helps the discovery of knowledge in large databases. A bibliometric study in this area is relevant for providing information and guidance to researchers regarding the progress of scientific production. This research aims to map the scientific production in the area of educational evaluation, focused on the PISA test using data mining techniques, presenting a broader view of national and international scientific production within this context. For this purpose, a bibliometric study of the scientific production indexed in the Scopus database was carried out considering the period of the last ten years. The final results point to the absence of studies that use data mining to make educational evaluations on the basis of PISA. This absence indicates a production deficit in this area, which has great potential to be explored.

Keywords: *Educational Indicators. PISA. Data Mining.*

3.3. Introdução

Considerando a complexidade do cenário mundial atual, repensar o ensino no país deve fazer parte do esforço para se criar condições de uma plena inserção no plano internacional. Isso sugere um grande desafio, que está relacionado aos esforços para se melhorar a eficiência das políticas públicas (CHESINI, 2014).

Os países em desenvolvimento precisam de resultados de grande escala para suas políticas educacionais e assim melhorar seu sistema educacional. Atendendo a necessidade de desenvolver evidências comparadas entre países com base no desempenho dos estudantes, em 1997, foi criado o Programa Internacional de Avaliação de Estudantes (PISA) pela Organização para a Cooperação e Desenvolvimento Econômico (OCDE).

Segundo o INEP (2016), a cada período do PISA, os questionários e um número de itens de cada área avaliada são disponibilizados pela OCDE para que educadores e pesquisadores compreendam melhor como os instrumentos foram construídos. Para Fonseca e Namen (2016) é evidente a exigência de um estudo aprofundado dos dados educacionais. Para o cumprimento de tal exigência, há a necessidade do emprego de tecnologias que possibilitem analisar uma vasta quantidade de dados e, por conseguinte, extrair informações e conhecimento relevante. A mineração de dados (*Data Mining* - DM) é uma dessas ferramentas, na qual auxilia a descoberta de conhecimento em grandes bases de dados, transformando dados em conhecimento útil de forma inteligente e automática (REIS, 2015).

Para Redondo *et al.* (2017), os estudos bibliométricos são ferramentas úteis para avaliar a relevância social e científica de um determinado assunto, uma vez que permitem a análise do crescimento e da distribuição na literatura científica. O objetivo dessa pesquisa é mapear a produção

científica na área de avaliação educacional com foco no PISA, utilizando técnicas de DM. Propõe-se identificar como se encontra esta temática de pesquisa no meio científico, e oferecer a potenciais pesquisadores da área, informações relevantes para estudos sobre esta matéria.

3.4. Metodologia

Foi realizado um estudo bibliométrico com o objetivo de identificar os trabalhos na área de avaliação educacional com foco no PISA, utilizando técnicas de DM, bem como de apresentar uma visão da produção científica nacional e internacional dentro deste contexto. Os procedimentos que foram executados nesse estudo corresponderam aos seguintes passos:

- definição da amostra;
- definição das palavras-chave e tesouros;
- pesquisa na base de dados selecionada;
- coleta de dados;
- organização dos resultados em diagrama específico;
- análise dos resultados.

A base de dados definida como amostra para a realização do estudo bibliométrico foi a *Scopus*. Segundo Redondo *et al.* (2017), trata-se do maior banco de dados de citação de literatura revisada por pares: revistas científicas, livros e conferências. Ela abrange cerca de 22 000 títulos de mais de 5 000 editores, dos quais 21 500 são revistas revisadas por pares nas ciências científicas, técnicas, médicas e sociais (incluindo artes e humanidades).

Nessa pesquisa, foram utilizados somente os artigos publicados em periódicos e conferências, tomando como fonte de dados as publicações dos últimos 10 anos (01/2008 a 09/2017). Foram definidas as palavras-chave: avaliação, educação, mineração de dados e PISA, e também os sinônimos e termos que as representam na literatura. Em seguida, foi realizado o levantamento na base *Scopus*. Por último, foi realizada a coleta de dados com o objetivo de realizar uma análise da evolução cronológica da produção científica, dos trabalhos e autores mais citados e das fontes de publicação.

3.5. Referencial Teórico

3.5.1. Estudo Bibliométrico

O conhecimento é resultante de informações interpretadas, um entendimento que surge da mistura de dados, interpretação individual, informação e experiência. O gerenciamento de conhecimento está aliado a ciência da informação, e é o meio a ser usado para identificar, capturar, avaliar, recuperar e compartilhar todos os recursos de informação. Para tanto, deve fazer uso de métodos, tecnologias e ferramentas almejando uma implementação bem-sucedida (UPADHYAY; KUMBHARANA, 2015).

Okubo (1997) definiu a bibliometria como uma ferramenta pela qual o estado da ciência e tecnologia pode ser observado através da produção geral de literatura científica. Ao fornecer novas informações, a bibliometria pode ser uma ajuda à tomada de decisões e ao gerenciamento de pesquisas. Segundo Alvarez e Caregnato (2017), o desenvolvimento dos estudos bibliométricos está relacionado com a evolução da ciência da informação.

Segundo Merigó e Yang (2017), a análise bibliométrica é uma técnica que está recebendo atenção crescente pela comunidade científica. Por isso, ela tem de ser tornada uma metodologia fundamental para realizar estudos com o objetivo de fornecer uma visão geral de um campo de pesquisa. Os dados fornecidos por meio dos estudos bibliométricos mensuram a contribuição do conhecimento científico proveniente das publicações em determinadas áreas.

Para Paula *et al.* (2017), a bibliometria pode ser utilizada para diversos fins. Por exemplo, pode-se verificar as tendências de publicações nas diversas áreas do conhecimento, acompanhar os estudos desenvolvidos nas diferentes esferas do conhecimento científico e, ainda, saber se o tema está antigo e deixou de ter determinada importância acadêmica. Para analisar a informação bibliográfica, é necessário selecionar a base de dados a ser usada.

A fonte de bibliometria é sempre um banco de dados. Várias bases estabelecidas por empresas ou por instituições públicas ou privadas são utilizadas para ilustrar os resultados da atividade científica e tecnológica. Após a realização de um processamento, as bases podem ser usadas para estabelecer indicadores bibliométricos (OKUBO, 1997).

3.5.2. Scopus

Scopus é uma base de dados internacional criada pela *Elsevier* em 2004, contendo informações bibliográficas de gênero científica e multidisciplinar. Ela contém mais de 53 milhões de referências publicadas em mais de 21 mil periódicos, 145 mil livros científicos 400 mil artigos de conferência e 27 milhões de patentes (MACHIN-MASTROMATTEO; TARANGO; MEDINA-YLLESCAS, 2017; SÁNCHEZ; DE LA CRUZ DEL RÍO RAMA; GARCÍA, 2017). Por vezes, em diversos lugares do mundo, outro repositório de dados utilizado é a *Web of Science* (WoS), que consta de uma base famosa no universo acadêmico (MACHIN-MASTROMATTEO; TARANGO; MEDINA-YLLESCAS, 2017).

3.5.3. Avaliação Educacional e o PISA

Nos últimos anos, algumas mudanças expressivas aconteceram na educação, incluindo os sistemas de avaliação e gestão educacional. Esse fato ocorreu em virtude de uma maior participação da sociedade nas políticas públicas em busca de melhorias e qualidade na educação. Isso contribuiu para o aumento da prática de avaliações com resultados divulgados por instituições de ensino (FERNANDES, 2016).

Segundo Gouw, Mota e Bizzo (2016), o objetivo da avaliação educacional é subsidiar o processo de ensino e aprendizagem, fornecendo informações sobre estudantes, professores e escolas, atuando como suporte para certificação, seleção e orientação de formulação de políticas educacionais. Existe uma série de avaliações nacionais e internacionais dos sistemas de ensino que permitem fazer o monitoramento do sistema educacional. O PISA é uma dessas avaliações, sendo um estudo realizado pela OCDE para avaliar os sistemas de educação de vários países e economias ao redor do mundo (MEDEIROS; JALOTO; SANTOS, 2017).

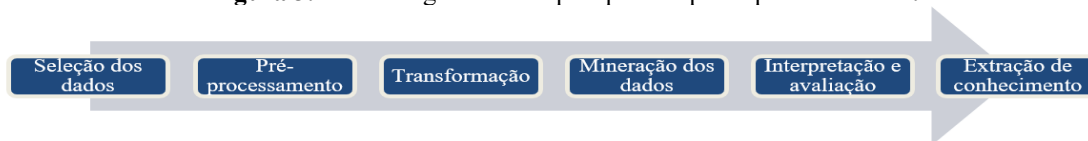
Segundo INEP (2016), o estudo proposto pelo PISA é feito a cada três anos e permite aos países participantes aferir conhecimentos e habilidades dos estudantes de 15 anos em Leitura, Matemática e Ciências, contrastando com resultados do desempenho de alunos dos países membros da OCDE, além de 35 países/economias parceiras. O PISA é um programa contínuo que, sob uma visão de longo prazo, tem por objetivo o desenvolvimento de um corpo de informações para o monitoramento de conhecimentos e habilidades dos estudantes em vários países, bem como em diferentes subgrupos demográficos de cada país (INEP, 2016).

Para Araújo e Tenório (2017), a análise dos resultados de uma avaliação é um elemento essencial para subsidiar as tomadas de decisões; é o elo entre o diagnóstico, a tomada de decisão e a melhoria do processo.

3.5.4. Mineração de Dados (DM)

Cardoso e Machado (2008) afirmaram que, mediante a existência de bancos de dados volumosos e com informações valiosas, faz-se necessária a aplicação de técnicas e ferramentas automáticas que viabilizem a extração de conhecimento de tais informações. Nesse cenário, surgiu o recurso de descoberto de conhecimento *knowledge discovery in databases* (KDD). Tal técnica foi definida por Fayyad, Piatetsky-Shapiro e Smyth (1996), como sendo um processo de várias etapas, não-trivial, interativo e iterativo, para a identificação de padrões compreensíveis, válidos, novos e potencialmente úteis a partir de grandes conjunto de dados. Tal processo foi mostrado na Figura 3.1.

Figura 3.1 – Visão geral das etapas que compõe o processo KDD.



Fonte: Paulista *et al.* (2017).

Wagner, Ribeiro e Zabadal (2016) definiram DM como um processo de extração de informações de uma grande base de dados para tomada de decisões, que é aplicada em diversas áreas como empresas, pesquisas e indústrias que utilizam os resultados, como exemplo, para melhoria de processos ou analisar tendências.

3.6. Resultados

Nesse trabalho, foi realizado um estudo bibliométrico na base de conhecimento *Scopus* com as palavras-chaves: avaliação, educação, mineração de dados e PISA. Além das palavras-chave mencionadas, foram definidos seus tesauros e termos, conforme mostrado no Quadro 3.1.

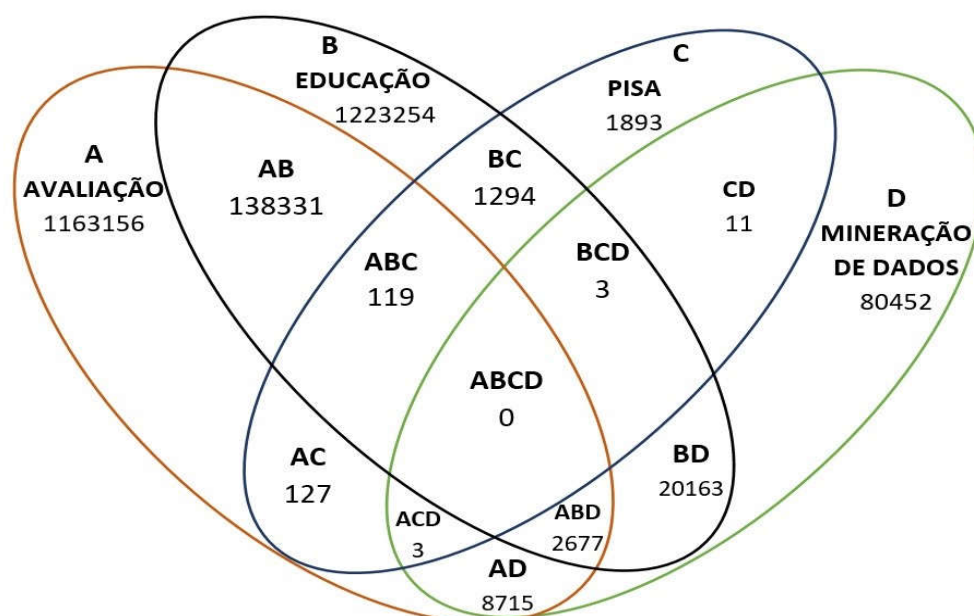
Quadro 3.1 – Termos utilizados para a pesquisa.

TITLE-ABS-KEY (evaluation OR assesment OR appraisal)	Tesaurus de A
TITLE-ABS-KEY (education OR teach* OR learning* OR school*)	Tesaurus de B
TITLE-ABS-KEY (pisa)	Tesaurus de C
TITLE-ABS-KEY (data AND mining*)	Tesaurus de D

Fonte: Elaboração própria (2017).

A Figura 3.2 apresenta o diagrama de *Venn* com a quantidade de trabalhos encontrados na base de conhecimento *Scopus*, considerando todas as intersecções possíveis entre os termos. Ressalta-se que a atualização do diagrama de *Venn* encontra-se no Apêndice A.

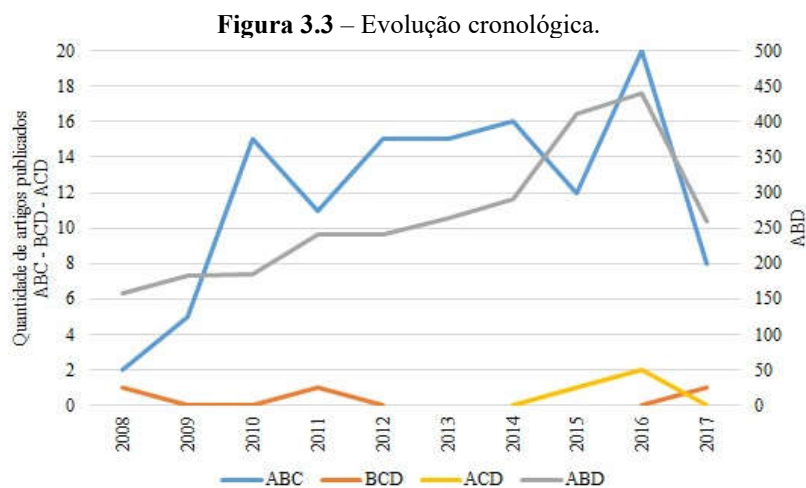
Figura 3.2 – Diagrama de *Venn* com a quantidade de trabalhos encontrados na base *Scopus*.



Fonte: Elaboração própria (2017).

Não foram encontrados trabalhos na intersecção dos quatro termos. Na intersecção ACD (avaliação, PISA e mineração de dados) e BCD (educação, PISA e mineração de dados) um número muito reduzido de artigos foram encontrados. Em ambos foram retornados 3 trabalhos, o que indica que existe um *déficit* de produção nessa área.

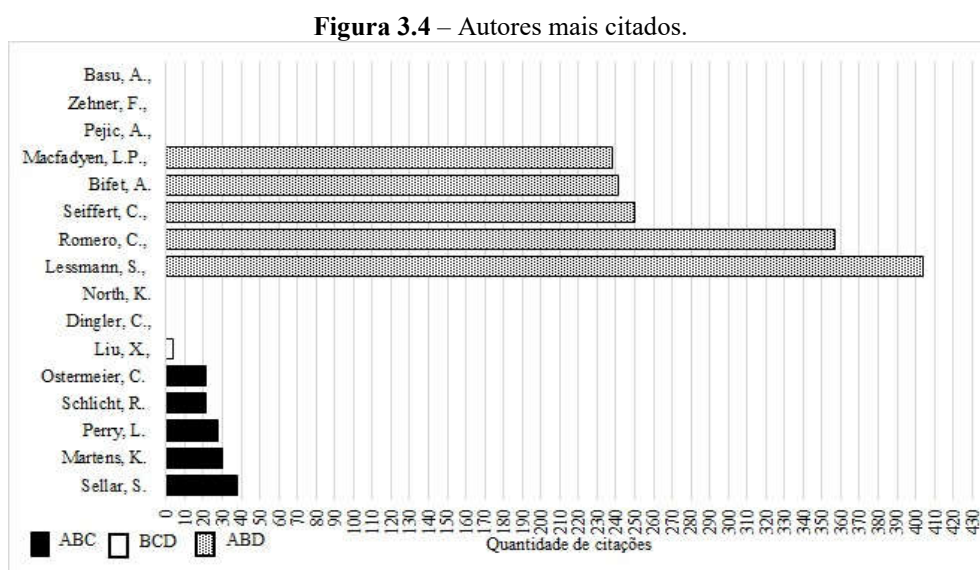
A Figura 3.3 apresenta a evolução cronológica das publicações encontradas de acordo com as áreas pesquisadas. A quantidade de trabalhos encontrada na intersecção ABD (avaliação, educação e mineração de dados) destoa das demais. Assim, os dados foram mapeados de forma que a quantidade de artigos publicados na área ABD está com ordenadas à direita e as outras intersecções (ABC, BCD e ACD) apresentam ordenadas à esquerda.



Fonte: Elaboração própria (2017).

Observou-se um crescimento da produção na área dos termos ABD e um pico de produção nas áreas dos termos ABC, ACD e ABD, em 2016. Pontes (2015) indicou que existe um interesse crescente na investigação e preposição de mecanismos e ferramentas voltados para o aprimoramento da avaliação educacional. Através do gráfico, pôde-se notar que as áreas que incluem os termos PISA e mineração de dados ainda não foram bem exploradas.

De posse do panorama da evolução da produção científica, foi efetuado o levantamento dos diversos autores que tiveram artigos publicados no período da pesquisa. A Figura 3.4 apresenta um mapeamento dos autores mais citados no período. Os autores da área dos termos ACD (Basu, Zehner e Pejic) e dois autores da área dos termos BCD (North e Dingler) possuem trabalhos publicados na área, mas não receberam citações.



Fonte: Elaboração própria (2017).

Foram identificados os autores que publicaram obras com alto impacto no universo científico. A área ABD se apresentou como a de maior interesse em relação às outras. Lessmann aparece como autor mais citado.

Os trabalhos mais citados entre os autores sobre a área pesquisada foram mapeados. Eles devem servir de base para a construção do referencial teórico de futuras pesquisas. O resultado dos 5 trabalhos mais citados para cada área é mostrado no Quadro 3.2. Foram listados apenas 3 artigos referentes as áreas representadas pelos termos BCD e ACD, tendo em vista que foram os únicos encontrados. As obras foram dispostas em ordem decrescente de citações.

Quadro 3.2 – Trabalhos mais citados.

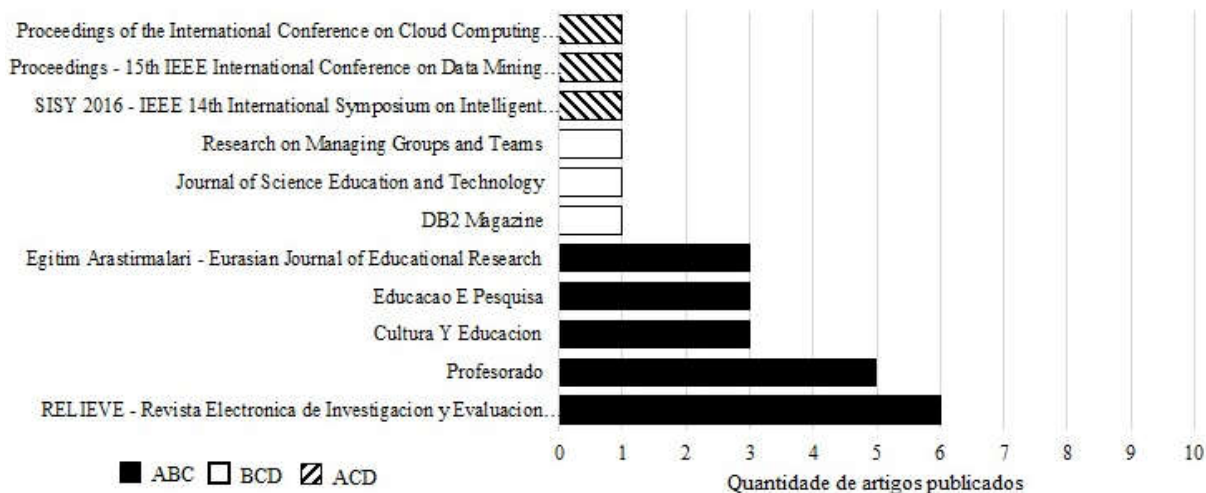
Termos	Nome do Artigo	Autores	Citações
ABC (Avaliação, Educação e PISA)	The OECD and the expansion of PISA: New global modes of governance in education	Sellar, S. Lingard, B.	38
	When do numbers count? The differential impact of the PISA rating and ranking on education policy in Germany and the US	Martens, K. Niemann, D.	30
	School socio-economic composition and student outcomes in Australia: Implications for educational policy	Perry, L. McConney, A	28
	Educational inequality in the EU: The effectiveness of the national education policy	Schlicht, R. Stadelmann-Steffen, I. Freitag, M.	21
	Improving science and mathematics instruction: The SINUS Project as an example for reform as teacher professional development	Ostermeier, C. Prenzel, M. Duit, R.	21
BCD (Educação, PISA e mineração de dados)	Opportunities-to-Learn at Home: Profiles of Students With and Without Reaching Science Proficiency	Liu, X. Whitford, M.	4
	Methodological challenges in measuring collaborative problem-solving skills over time	Dingler, C. Von Davier, A. Hao, J.	0
	XML: Powering next-generation business applications	North, K.	0
ABD (Avaliação, educação e mineração de dados)	Benchmarking classification models for software defect prediction: A proposed framework and novel findings	Lessmann, S. Baesens, B. Mues, C. Pietsch, S.	404
	Data mining in course management systems: Moodle case study and tutorial	Romero, C. Ventura, S. García, E.	357
	RUSBoost: A hybrid approach to alleviating class imbalance	Seiffert, C. Khoshgoftaar, T.M. Van Hulse, J. Napolitano, A.	250
	New ensemble methods for evolving data streams	Bifet, A. Holmes, G.	241

		Pfahringher, B. Kirkby, R. Gavaldà, R.	
	Mining LMS data to develop an "early warning system" for educators: A proof of concept	Macfadyen, L.P. Dawson, S.	238
ACD (Avaliação, PISA e mineração de dados)	Exploring data mining possibilities on computer based problem solving data	Pejic, A. Molcer, P.S.	0
	Using and Improving Coding Guides for and by Automatic Coding of PISA Short Text Responses	Zehner, F. Goldhammer, F. Salzer, C.	0
	CF-inspired privacy-preserving prediction of next location in the cloud	Basu, A. Corena, J.C. Monreale, A. Vaidya, J. Miyake, Y.	0

Fonte: Elaboração própria (2017).

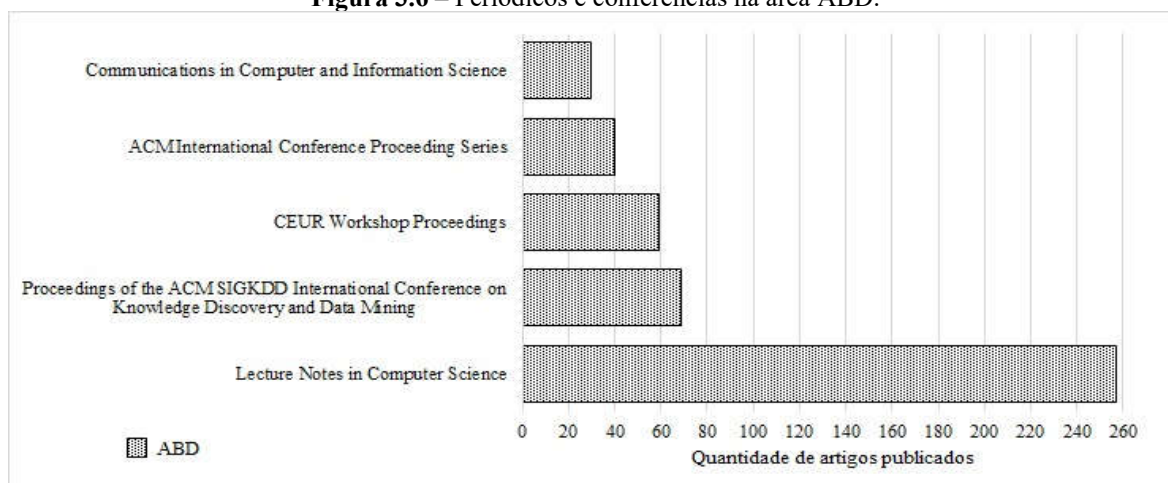
As Figuras 3.5 e 3.6 apresentam cinco periódicos e/ou conferências que tiveram maior número de publicação dentro da respectiva área de acordo com termos pesquisados. Tendo em vista que a quantidade de publicações na área dos termos ABD se destacou das outras, foi necessário dividir essa análise em dois gráficos, para melhor visualização dos dados.

Figura 3.5 – Periódicos e conferências nas áreas ABC, BCD e ACD.



Fonte: Elaboração própria (2017).

Figura 3.6 – Periódicos e conferências na área ABD.



Fonte: Elaboração própria (2017).

De acordo com a Figura 3.6, um número elevado de artigos foi publicado pela *Lecture Notes in Computer Science*, o que demonstrou que esse periódico tem um importante papel na área e deve ser acompanhado para pesquisas e futuras publicações. Tal revista foi avaliada pela CAPES em 21 áreas, onde se destacaram as melhores avaliadas: administração pública e de empresas, ciências contábeis e turismo (A2); planejamento urbano e regional/demografia (A2); e interdisciplinar (B1).

3.7. Conclusão

O objetivo principal desta pesquisa foi mapear a produção científica na área de avaliação educacional com foco no Programa Internacional de Avaliação de Estudantes (PISA), utilizando técnicas de DM, a partir de artigos indexados no *Scopus*. Os resultados do presente estudo favoreceram ao conhecimento de um cenário mais abrangente sobre a produção científica sobre o tema. Foi verificado que a área tem um grande potencial a ser explorado, tendo em vista a inexistência de artigos publicados numa pesquisa envolvendo os termos e sinônimos de avaliação, educação, PISA e mineração de dados.

Foi possível verificar por meio da evolução cronológica das publicações encontradas, que a área de avaliação educacional usando DM se destacou em termos de quantidade de publicações e apresentou uma produção crescente ao longo dos anos estudados. Analisando os autores mais citados, foi possível identificar que os trabalhos publicados pelos autores da área dos termos ACD e 2 autores da área dos termos BCD não receberam citações. Dentre todas as intersecções, o autor mais citado é o Lessmann, S., com trabalhos publicados envolvendo avaliação educacional utilizando DM (ABD).

O mapeamento dos trabalhos mais citados permitiu identificar os trabalhos clássicos que poderá servir de apoio para pesquisadores e sustentação de novos trabalhos no processo de revisão de literatura. Para trabalhos futuros, sugere-se a ampliação desse estudo utilizando outras bases de dados de produção científica. Além da base *Scopus*, é senso comum a utilização da base *ISI – Web of Science*. Os termos e sinônimos podem ser ampliados com utilização de ferramentas de suporte ao estudo bibliométrico. Dentre elas, pode-se citar o *Sitakis and Ucinet*. Propõe-se, por fim, a elaboração de um material científico que demonstre o estado da arte da pesquisa proposta. Tal trabalho pode auxiliar os pesquisadores a se situarem no atual estado em que se encontra os estudos acerca do escopo de pesquisa selecionado para análise.

Referências Bibliográficas

ALVAREZ, G. R.; CAREGNATO, S. E. A ciência da informação e sua contribuição para a avaliação do conhecimento científico. **Biblos**, v. 31, n. 1, p. 09–26, 2017.

ARAÚJO, M. DE L. H. S.; TENÓRIO, R. M. Resultados brasileiros no PISA e seus (des)usos. **Estudos em Avaliação Educacional**, v. 28, n. 68, p. 344, 31 ago. 2017.

CARDOSO, O. N. P.; MACHADO, R. T. M. Gestão do conhecimento usando data mining: estudo de caso na Universidade Federal de Lavras. **Revista de administração pública**, v. 42, n. 3, p. 495–528, jun. 2008.

CHESINI, T. S. **Aplicação de técnicas de geoprocessamento para subsidiar a análise e tomada de decisão no âmbito da educação**. Porto Alegre/RS: Pontifícia Universidade Católica do Rio Grande do Sul, 2014.

FAYYAD, U.; PIATETSKY-SHAPIRO, G.; SMYTH, P. From data mining to knowledge discovery in databases. **AI magazine**, v. 17, n. 3, p. 37, 1996.

FERNANDES, R. A universalização da avaliação e a criação do Ideb: pressupostos e perspectivas. **Em Aberto**, v. 29, n. 96, 2016.

FONSECA, S. O. DA; NAMEN, A. A. Data Mining On Inep Databases: an Initial Analysis Aiming to Improve Brazilian Educational System. **Educação em Revista**, v. 32, n. 1, p. 133–157, mar. 2016.

GOUW, A. M. S.; MOTA, H. S.; BIZZO, N. M. V. Brazilian Youth and Science: Possible Relationship of Interest. **Revista Brasileira de Pesquisa em Educação em Ciências**, v. 16, n. 3, p. 649–670, 2016.

INEP. **Brasil no PISA 2015 : análises e reflexões sobre o desempenho dos estudantes brasileiros / OCDE-Organização para a Cooperação e Desenvolvimento Econômico**. Brasília/DF: Fundação Santillana, 2016.

MACHIN-MASTROMATTEO, J. D.; TARANGO, J.; MEDINA-YLLESCAS, E. Latin American triple-A journals 1: A quality roadmap from the quality indicators and journals' presence in Web of Science and Scopus. **Information Development**, v. 33, n. 4, p. 436–441, set. 2017.

MEDEIROS, L.; JALOTO, A.; SANTOS, A. V. F. DOS. A área de ciências nas avaliações internacionais de larga escala. **Estudos em Avaliação Educacional**, v. 28, n. 68, p. 512, 31 ago. 2017.

MERIGÓ, J. M.; YANG, J.-B. A bibliometric analysis of operations research and management science. **Omega**, v. 73, p. 37–48, dez. 2017.

OKUBO, Y. **Bibliometric Indicators and Analysis of Research Systems**, 1997. Disponível em: <http://www.oecd-ilibrary.org/science-and-technology/bibliometric-indicators-and-analysis-of-research-systems_208277770603>. Acesso em: 24 out. 2018.

PAULA, R. S. P. DE *et al.* Indicadores bibliométricos na base scopus: Uma análise das publicações sobre o tema “economia ambiental”. **Brazilian Journal of Development**, v. 3, n. 2, p. 350–365, 2017.

PAULISTA, C. R. *et al.* **Análise do comportamento de variáveis climáticas sob a ótica da geração energética renovável**. Proceedings CLME2017/VCEM. **Anais...** In: 8º Congresso Luso-Moçambicano de Engenharia / V Congresso de Engenharia de Moçambique. Maputo, Moçambique: INEGI/FEUP, 2017.

PONTES, L. A. F. **Medidas de eficácia escolar no contexto das políticas brasileiras de responsabilização educacional: o caso do Índice de Desenvolvimento da Educação Básica, o IDEB, em Minas Gerais**. Dissertação (Doutorado em Educação) - Juiz de Fora/MG: Universidade Federal de Juiz de Fora, 2015.

REDONDO, M. *et al.* A bibliometric study of the scientific publications on patient-reported outcomes in rheumatology. **Seminars in Arthritis and Rheumatism**, v. 46, n. 6, p. 828–833, 2017.

REIS, R. X. DOS. **A utilização do fenômeno Big Data na Administração Pública - a experiência do PENSA na Prefeitura Municipal do Rio de Janeiro**. Trabalho de conclusão de curso—Rio de Janeiro: Escola de Contas e Gestão, 2015.

SÁNCHEZ, A. D.; DE LA CRUZ DEL RÍO RAMA, M.; GARCÍA, J. Á. Bibliometric analysis of publications on wine tourism in the databases Scopus and WoS. **European Research on Management and Business Economics**, v. 23, n. 1, p. 8–15, jan. 2018.

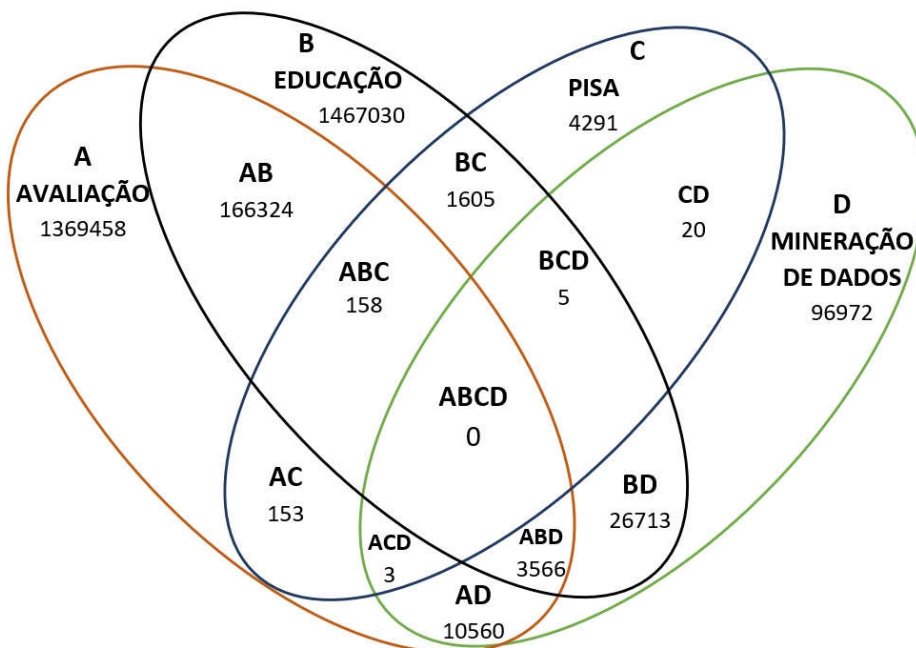
UPADHYAY, A.; KUMBHARANA, C. K. **Analysis of functional parameters to implement knowledge management for sustainable e-governance in agriculture sector of saurashtra region of gujarat state**. Advances in Intelligent Systems and Computing. **Anais...** In: 50TH ANNUAL CONVENTION OF COMPUTER SOCIETY OF INDIA : ICT BASED INNOVATIONS, CSI 2015. 2015. Acesso em: 23 out. 2018

WAGNER, M. M.; RIBEIRO, V.; ZABADAL, J. Mineração de Dados Meteorológicos Empregando Dados de Temperatura: o caso de uma cidade gaúcha. **Interfaces Científicas - Exatas e Tecnológicas**, v. 2, n. 2, p. 57–70, 11 out. 2018.

Apêndice A

A Figura 3.7 apresenta o diagrama de *Venn* atualizado em dezembro de 2018 com a quantidade de trabalhos encontrados na base de conhecimento *Scopus*, considerando todas as intersecções possíveis entre os termos.

Figura 3.7 – Diagrama de *Venn* atualizado com a quantidade de trabalhos encontrados na base *Scopus*.



Fonte: Elaboração própria (2018).

Verificou-se que na intersecção dos quatro termos (ABCD) a quantidade de trabalhos encontrados ainda é nula. As intersecções ACD (avaliação, PISA e mineração de dados) e BCD (educação, PISA e mineração de dados) continuam com um número muito reduzido de artigos encontrados, mostrando assim que permanece um *déficit* de produção nessa área. Na área BCD (educação, PISA e mineração de dados) foram encontrados mais 2 trabalhos, um de Chang e Chen (2018) e o outro de Qiao e Jiao (2018), que serviram de base para pesquisa metodológica, apresentada na Seção 4.4.3, do Capítulo 4.

4. ARTIGO II: IDENTIFICAÇÃO DOS ATRIBUTOS DISCENTES RELEVANTES NO RESULTADO DO PISA UTILIZANDO TÉCNICAS DE MINERAÇÃO DE DADOS

4.1. Resumo

A exigência de um estudo aprofundado dos dados educacionais pressupõe o emprego de tecnologias que possibilitem analisar uma vasta quantidade de dados e, por conseguinte, extrair informações e conhecimento relevante. O objetivo do presente trabalho é aplicar técnicas de mineração de dados na base de dados do PISA para identificar padrões em potencial que expliquem o sucesso dos países que se destacaram na avaliação. Na metodologia foram realizadas as etapas de obtenção dos dados, criação do banco e extração dos dados dos países, pré-processamento e higienização dos dados e mineração de dados. Na última etapa foi utilizado o método de classificação J48. A partir das árvores de decisão obtidas foi possível identificar os atributos que foram relevantes. Esses atributos estão relacionados aos seguintes temas: aspiração do nível educacional do aluno, reprovação, motivação e ansiedade, fatores socioeconômicos, abordagens científicas, uso das tecnologias de informação e comunicação, interações com amigos, prática de atividade de física, trabalho remunerado, tarefa de casa, tempo de aprendizagem em cada disciplina, cooperação e trabalho em equipe, programa de estudo do aluno, imparcialidade do professor e série/ano do aluno. Sendo assim, os resultados foram considerados satisfatórios por terem permitido realizar a análise dos atributos que foram relevantes nos países que se destacaram na avaliação.

Palavras-chave: Avaliação; Educação; PISA; Indicadores educacionais.

4.2. Abstract

The requirement of a detailed study of educational data presupposes the use of technologies that make it possible to analyze a vast amount of data and extract relevant information and knowledge. This research aims to apply data mining techniques in the PISA database to identify potential patterns that explain the success of the countries that excelled in the assessment. In the methodology were done: the steps of data collection, bank creation and country data extraction, data pre-processing and hygiene and data mining. In the last step, J48 classification method was applied. Through the obtained decision trees, it was possible to identify the relevant attributes. These attributes are related to the following topics: aspiration of student's educational level, disapproval, motivation and anxiety, socioeconomic factors, scientific approaches, use of information and communication technologies, interactions with friends, practice of physics activity, paid work, homework, learning time by discipline, cooperation and teamwork, student study program, teacher impartiality and student grade. Therefore, the results were satisfactory because they provided an analysis of the attributes that were relevant to the countries that were highlighted in the evaluation.

Keywords: Evaluation; Education; PISA; Educational indicators

4.3.Introdução

Nas últimas décadas, o debate de qualidade em educação foi além do contexto educacional, ganhando grande visibilidade e se tornando a preocupação dos especialistas de diversas áreas de conhecimento (MARQUES, 2017). Para Oliveira, Coelho e Castanha *et al.* (2015), a discussão dos problemas da educação básica e os desafios para melhorar a sua qualidade têm sido caracterizados pela divulgação ampla de informações produzidas pelos sistemas de avaliação, focada no rendimento do aluno e no desempenho dos sistemas de ensino.

Segundo Badri *et al.* (2016), a avaliação educacional é importante não só para avaliar a qualidade das escolas, mas também para identificar os seus pontos fortes e fracos, procurando garantir que todas as instituições ofereçam aos alunos uma educação de qualidade que atenda aos padrões de excelência alcançados por países que são referências em educação.

Moura e Sadoyama (2017) afirmaram que além de exercer seu papel de facilitador e promotor do desenvolvimento pleno do ser, as instituições escolares devem estar preparadas para prestar serviços educacionais de qualidade. Isso se dá, pois estão diretamente ligadas com a formação do sujeito.

No campo educacional, Salah e Saadi (2017) consideraram a qualidade de serviço como o conjunto de propriedades quantitativas e qualitativas que descrevem como funciona um serviço e que podem atender às necessidades do aluno. Para aferir essa qualidade, faz-se necessário o uso de indicadores. Para Fernandes (2016), é necessário produzir informações e indicadores que possibilitem à sociedade julgar a qualidade do serviço educacional que é oferecido.

Segundo Muñoz e Charro (2018), para gerar indicadores de desempenho que permitam definir políticas de longo alcance com base nas fraquezas e pontos fortes dos sistemas educacionais, foi criado o Programa Internacional de Avaliação de Estudantes (PISA) pela Organização para a Cooperação e Desenvolvimento Econômico (OCDE). O estudo proposto pelo PISA é feito a cada três anos e permite aos países participantes aferir conhecimentos e habilidades dos estudantes em Matemática, Leitura e Ciências, contrastando com resultados do desempenho de alunos dos países membros da OCDE, além dos países/economias parceiras (INEP, 2016).

Ciceo (2018) ressaltou que os testes do PISA são uma fonte confiável de informações para monitorar os resultados dos sistemas educacionais dentro de um quadro internacional, e uma ferramenta útil para refinar as políticas educacionais. Gamazo *et al.* (2018) afirmaram que o acesso

aberto dos resultados do PISA facilita o estudo do impacto dos fatores cultural, econômico, social e educacional no desempenho dos alunos.

Com uma adequada análise de dados, os fatores que afetam o desempenho de um aluno podem ser descobertos. Nesse contexto, surge a mineração de dados educacionais, que é uma área emergente da mineração de dados (*Data Mining* - DM), na qual técnicas são aplicadas a um volume de dados com o objetivo de reconhecer indicadores importantes para avaliar o sistema educacional e ter informações sobre a aprendizagem dos alunos (KASTHURIARACHCHI; LIYANAGE; BHATT, 2018).

Liu e Whitford (2011) utilizaram DM para criar padrões de associação entre as variáveis de oportunidade de aprender em casa e a obtenção de informações de proficiência científica em uma amostra de 2006 do PISA. Verificou-se que os estudantes que não atingiram a proficiência científica são caracterizados por terem menos de 100 livros em casa. Por outro lado, os estudantes que atingiram a proficiência em Ciências se caracterizam por ter mais de 100 livros em casa e terem um nível mais alto de educação. Além das características apontadas acima, são também identificadas outras características do domicílio (computador e *internet* em casa e língua falada em casa) nos perfis dos alunos que atingiram a proficiência em Ciências.

Avsar e Tavsancil (2014) realizaram um estudo para determinar as variáveis que explicam a alfabetização de estudantes turcos, por meio de sua resposta ao questionário PISA 2009, a partir do questionário do aluno PISA, utilizando a análise CHAID (*Chi-squared automatic interaction detection*). Trata-se de um dos algoritmos de árvores de decisão de mineração de dados. O resultado do estudo indicou que a variável que melhor definiu o desempenho dos estudantes turcos em Leitura foi a posse de livros literários.

Aksu e Guzeller (2016) realizaram um estudo das variáveis relacionadas a disciplina e os efeitos na classificação da Turquia no PISA, para classificar os alunos bem-sucedidos e insucessos em termos de alfabetização Matemática. Na etapa de DM foi utilizada a análise CHAID. A percepção de auto eficácia, a atitude em relação ao curso e disciplina de estudo foram as características afetivas mais importantes na classificação dos alunos bem e malsucedidos.

Até o momento, no entanto, não foram encontrados estudos que tenham utilizado técnicas de DM para analisar os fatores que contribuem para o desempenho dos alunos na avaliação do PISA. Este trabalho visa preencher essa lacuna e fornece uma didática do processo de análise dos dados do PISA de 2015. Para tanto, o objetivo da presente pesquisa é aplicar técnicas de DM na base de dados

do PISA para identificar padrões em potencial que expliquem o sucesso dos países que se destacaram na avaliação.

4.4. Metodologia

4.4.1. Classificação da pesquisa

A classificação da presente pesquisa foi definida de acordo com 3 pontos de vista. Quanto a natureza, a pesquisa é aplicada pois visa gerar informações para empregar na resolução de problemas educacionais, seja ele no contexto local ou internacional; ela é quantitativa pois trata as informações de forma numérica, com análises dos resultados; a dissertação é descritiva, pois utiliza de técnicas para descrever os atributos e estabelece relações entre as variáveis (GIL, 2009).

4.4.2. População e amostra

A base de dados a ser utilizada para a aplicação das técnicas de DM foi a do PISA. Uma vez que os resultados são disponibilizados em frequência trienal, a presente pesquisa tratou os dados referentes ao ano de 2015. A base utilizada englobou dados dos 35 países da OCDE e 37 países parceiros e economias que participaram da avaliação. A amostra de alunos do PISA 2015 foi selecionada usando princípios reconhecidos profissionalmente de amostragem científica de forma a assegurar a representação do alvo total da população de estudantes de 15 anos nos países participantes (OCDE, 2017a).

Foram extraídos os dados referentes aos 5 países com maior classificação em cada área de conhecimento (Matemática, Leitura e Ciências), de acordo com a OCDE (2016a). A Tabela 4.1 mostra os países e a classificação na área de conhecimento no PISA 2015 juntamente com a quantidade de estudantes que fizeram a avaliação em cada país, que representa a amostra da pesquisa.

Tabela 4.1 – Amostra utilizada na pesquisa.

Países	Área de conhecimento	Classificação na área no PISA 2015	Número de estudantes
Singapura	Matemática	1º	6115
	Leitura	1º	
	Ciências	1º	
Hong Kong (China)	Matemática	2º	5359
	Leitura	2º	
Japão	Ciências	2º	6647
	Matemática	5º	

Macau (China)	Matemática	3º	4476
Canadá	Leitura	3º	20058
Estônia	Ciências	3º	5587
Taipei Chinesa	Matemática	4º	7708
	Ciências	4º	
Finlândia	Leitura	4º	5882
	Ciências	5º	
Irlanda	Leitura	5º	5741

Fonte: Elaboração própria (2018).

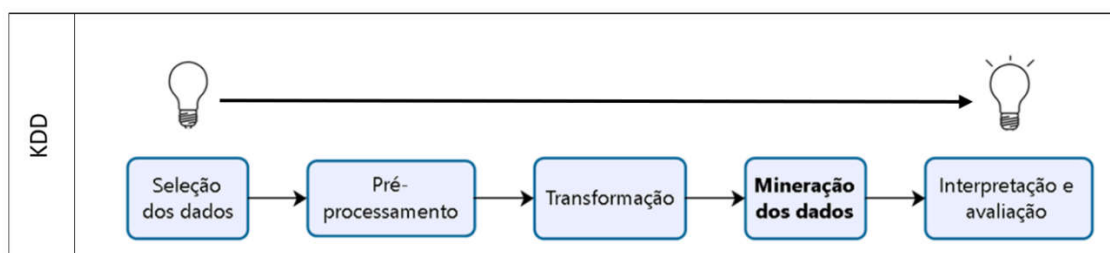
A base de dados é dividida em linhas e colunas, sendo composta de 920 atributos. Cada atributo ocupa uma coluna e cada linha corresponde a resposta de um aluno. Assim, por exemplo, como em Singapura 6.115 estudantes participaram da avaliação, conforme mostra a Tabela 4.1, tem-se que a base de dados desse país tem 5.625.800 dados.

4.4.3. Pesquisa metodológica

Para atender à necessidade urgente de converter um elevado volume de dados em informações e conhecimentos úteis, surgiu a técnica de DM. Essa área tem atraído grande importância na indústria da informação e na sociedade nos últimos anos (HUSSAIN *et al.*, 2018).

A DM faz parte de uma das etapas do processo de descoberta de conhecimento denominado *knowledge discovery in databases* (KDD). O KDD foi definido por Fayyad, Piatetsky-Shapiro e Smyth (1996), como sendo um processo de várias etapas, não-trivial, interativo e iterativo, para a identificação de padrões compreensíveis, válidos, novos e potencialmente úteis a partir de um grande volume de dados. Os autores apontaram que para extrair conhecimento dos dados é necessário seguir as seguintes etapas: seleção de dados, pré-processamento, transformação, mineração de dados e interpretação e avaliação. A Figura 4.1 mostra as etapas do KDD, de forma esquemática. Diversas técnicas podem ser utilizadas, a depender do objetivo a ser alcançado.

Figura 4.1 – Visão geral das etapas que compõe o processo KDD.



Fonte: Elaboração própria (2018).

Dentro da DM, a mineração de dados educacionais (DME) se apresenta como uma área emergente. A qualidade da educação precisa ser melhorada e a DME é uma ferramenta para essa melhoria. O campo educacional usa ferramentas de mineração de dados para explorar e analisar o desempenho dos alunos e avaliar melhor o processo de ensino e aprendizagem (HUSSAIN *et al.*, 2018). Com o objetivo de embasar a presente pesquisa e as etapas da metodologia proposta foram pesquisados trabalhos que utilizam técnicas de DM para análise de dados educacionais.

Ferreira (2015) utilizou a DME para estudar quais características individuais, familiares e locais contribuem para que os alunos possuam ou não o Ensino Fundamental concluído, focando nos alunos de 16 anos das escolas da cidade de Porto Alegre. A base de dados utilizado foi a do Censo Escolar da Educação Básica do Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (INEP) de 2014. A base de dados com as informações das matrículas escolares é composta de 85 variáveis, a das escolas é composta de 143 variáveis e a das turmas escolares por 79 variáveis. Foi realizada a agregação das três bases de dados em uma única, sendo criada a variável de interesse que identifica se o aluno possui ou não o Ensino Fundamental Completo. Os dados foram interpretados utilizando o *software* SPSS. Foram utilizados os métodos de mineração de dados J48 e o filtro *CfsSubsetEval*, ambos disponíveis no *software* WEKA. Os resultados apontaram que a precisão média obtida da árvore de decisão resultante da aplicação do filtro sobre os atributos da base de dados foi de 96,17%, gerando uma árvore com 29 nodos e associando recursos de *internet* banda larga, laboratório de ciências, auditório na escola e ensino privado com maiores chances de o aluno apresentar o Ensino Fundamental concluído. Foram encontradas evidências de que necessidades especiais dos alunos estejam ligadas com a não conclusão do Ensino Fundamental e de que a cor/raça branca, aulas de inglês, espanhol, artes e outras disciplinas não obrigatórias estejam associadas à conclusão do Ensino Fundamental.

Simon e Cazella (2017) empregaram a DME no banco de dados públicos referentes ao Exame Nacional do Ensino Médio (ENEM) de 2015. O objetivo foi de gerar um modelo preditivo do indicador de desempenho médio em ciências da natureza e suas tecnologias dos alunos de escolas do ensino médio. As variáveis selecionadas para pesquisa foram: dependência administrativa, localização, número de alunos no censo, taxa de participação, indicador de permanência na escola, indicador de nível socioeconômico, indicador de formação docente, taxa de aprovação e média da escola. O método de DM utilizado foi a árvore de decisão, através do algoritmo J48, com *software* WEKA. Nos resultados obtidos sobre o desempenho médio dos alunos em ciências da natureza e suas tecnologias, destaca-se que os valores acima de 550 pontos, ocorrem nas escolas: privadas, apenas no

nível socioeconômico muito alto; federais, nos níveis muito alto, alto e médio alto; estaduais, apenas no nível muito alto; e nas municipais, apenas no nível médio alto.

Baby (2018) realizou um levantamento de informações em relação aos professores, por meio de um questionário de avaliação aplicado aos alunos. O principal objetivo da pesquisa foi demonstrar a utilização de técnicas de DM para analisar o desempenho de professores. Os autores compararam o J48 com diversos outros algoritmos, e concluíram que o J48 é o método de classificação mais adequado para a predição de desempenho pedagógico.

Chang e Chen (2018) aplicaram o método de análise de *cluster* e o modelo de regressão *stepwise* para determinar o principal fator de impacto nos resultados do PISA de 2015 dentre alguns selecionados. Para tanto, o estudo utilizou os fatores referentes aos gastos do governo com educação, horas de ensino por ano e análise docente do ensino primário e secundário. O resultado revelou apenas que as horas de ensino por ano tem relação negativa com o desempenho do PISA de 2015.

Kasthuriarachchi *et al.*(2018) utilizaram a DME para encontrar os principais fatores que afetam o desempenho dos alunos matriculados em um programa de licenciamento em tecnologia em Sri Lanka. O conjunto de dados selecionado para este estudo consistia em 13 atributos, sendo eles: sexo, idade, *fjob* (se o pai do aluno trabalha), *failure* (número de módulos de falha), *extrapayment* (pagamento da mensalidade extra), *moreyears* (se o aluno ficou mais de 3 anos), *ispass* (se o aluno passou o diploma) e mais 6 atributos que estão relacionados ao desempenho em cada semestre que o aluno cursou (S1, S2, S3, S4, S5 e S6). Três métodos de DM foram usados para modelar o desempenho, que foram: algoritmo *Naive Bayes*, o algoritmo *Tree Decision* (C 5.0) e algoritmo *Random Forest* (RF). A idade do aluno, o número de módulos de falha e o desempenho nos semestres foram identificados como os fatores mais correlacionados que preveem a nota final dos alunos.

Nicoletti, Marques e Guimarães (2018) realizaram alguns experimentos com dados educacionais extraídos de registros de uma universidade, aplicando o algoritmo J48. O objetivo foi discutir técnicas para se prever o desempenho dos estudantes. Foi constatada a viabilidade da utilização dessa ferramenta, gerando inclusive uma possibilidade em potencial de aplicação durante o período, de forma a direcionar os esforços em relação aos alunos previstos para falhar.

Qiao e Jiao (2018) realizaram um estudo na base de dados do PISA de 2012 para comparar os métodos de mineração de dados de árvore de classificação CART, gradiente de reforço, floresta aleatória e SVM (máquina de vetores de suporte). A pesquisa teve a finalidade de se atingir 3 objetivos, que foram: (1) demonstrar o uso de métodos de mineração de dados para processar os dados

de forma sistemática; (2) avaliar a consistência da classificação das diferentes técnicas de mineração de dados, seja supervisionado ou não supervisionado; (3) analisar como os resultados podem ser usados para lidar com resolução de problemas. Todos os quatro métodos alcançaram uma precisão de classificação satisfatória com base em várias medidas de resultados, mostrando precisão geral e valores ligeiramente melhores. Em geral, a fácil interpretação e visualização gráfica são as principais vantagens das árvores.

De forma geral, os trabalhos relatados seguem as etapas do KDD, e na etapa da DM o método mais utilizado é o algoritmo J48.

4.4.4. Procedimentos técnicos

Os procedimentos que foram executados nessa pesquisa tiveram por base as etapas do KDD apresentadas na Figura 4.1 e corresponderam aos passos descritos na Figura 4.2.

Figura 4.2 – Etapas da metodologia da pesquisa.



Fonte: Elaboração própria (2018).

1) Obtenção dos dados

Os dados foram obtidos na base de dados do PISA de 2015, disponível em <http://www.oecd.org/pisa/data/2015database/>. A base utilizada foi a “*Student questionnaire data file*”, que tem um tamanho de 411 MB. Os arquivos de dados do PISA são de uso público e combinam dados dos 72 países participantes, relatados em um arquivo, que incluem um conjunto de variáveis internacionais. Para a conversão da base em arquivo de extensão .csv foi utilizado como auxílio o *software* SAS, do *SAS Institute Inc*, que é um sistema integrado de aplicações para a análise de dados.

2) Criação do banco e extração dos dados dos países

Devido ao tamanho da base de dados foi necessário realizar a conversão para uma estrutura de dados em *MySQL*. Primeiramente, foram necessários três *scripts*, em linguagem *Python*, para a criação do banco de dados, que são:

- *get_fields.py*: usado para extrair os nomes dos campos do arquivo;
- *create_table.py*: usado para fazer o código de criação do banco;
- *insert_into.py*: usado para fazer o código de inserção dos valores no banco.

O primeiro *script* faz uso somente da segunda linha do arquivo da base de dados, logo, não foi necessária nenhuma ação especial para o mesmo. Além disso, ele gera um arquivo com os nomes dos campos, usado pelos próximos *scripts*. O segundo *script* faz uso somente do arquivo gerado pelo primeiro, enquanto o terceiro *script* faz uso do arquivo gerado pelo primeiro *script* e do arquivo da base de dados. Após copiar o arquivo da base de dados e usar o comando *split -l 1000*, diversos arquivos foram gerados, cada um, contendo no máximo 1000 linhas. Depois de executar os *scripts* *create_table.py* (criação do banco de dados) e *insert_into.py* (insere valores no banco de dados), dois novos arquivos foram gerados. Para execução dos arquivos, foi utilizada a linha de comando do *MySQL* 10.1.36, com o comando *mysql -u root*. Para finalizar a criação do banco de dados, utilizou-se os comandos *source create_table.sql* e *source insert_into.sql*.

O *MySQL* é capaz de exportar os resultados de uma consulta *sql* diretamente para um arquivo. Assim, para extrair os dados dos 5 países com maior classificação em cada área de conhecimento (Matemática, Leitura e Ciências) foram utilizados os seguintes comandos: seleção dos campos desejados (*select campos_desejados*), código do país desejado (*where cnt = codigo_do_país_desejado*) e nome do arquivo destino (*into outfile nome_arquivo_destino*). Foram gerados 15 arquivos (5 países de cada área de conhecimento), sendo um por vez, no mesmo padrão do arquivo da base de dados.

3) Pré-processamento e higienização dos dados

Na terceira etapa foram realizados o pré-processamento e a higienização dos dados. Para tanto, foram realizados os seguintes procedimentos para cada área de conhecimento separadamente:

- a) Cálculo da média das notas em cada área de conhecimento. Essas notas são representadas pelos seguintes atributos:

- Matemática: PV1MATH, PV2MATH, PV3MATH, PV4MATH, PV5MATH, PV6MATH, PV7MATH, PV8MATH, PV9MATH e PV10MATH;
 - Leitura: PV1READ, PV2READ, PV3READ, PV4READ, PV5READ, PV6READ, PV7READ, PV8READ, PV9READ e PV10READ;
 - Ciências: PV1SCIE, PV2SCIE, PV3SCIE, PV4SCIE, PV5SCIE, PV6SCIE, PV7SCIE, PV8SCIE, PV9SCIE e PV10SCIE.
- b) Após o cálculo da média, exclusão dos atributos referentes as notas em área de conhecimento;
 - c) Divisão das médias em quartis, ficando as menores médias representadas pelo quartil 1(Q1) e as maiores médias pelo quartil 4 (Q4);
 - d) Exclusão dos atributos específicos das outras áreas;
 - e) Exclusão dos atributos com dados faltantes.

4) Mineração de dados

O *software* de DM utilizado foi o WEKA (*Waikato Environment for Knowledge Analysis*), na versão 3.8.1. O WEKA é uma ferramenta de código aberto escrita em Java que é amplamente utilizada na mineração de dados (HUSSAIN *et al.*, 2018). Através do *software* WEKA, foram geradas as árvores de decisão utilizando o algoritmo de classificação J48. Tal algoritmo é baseado no algoritmo C4.5 proposto por Quinlan (1987) e induz uma árvore de decisão a partir de uma base de dados em que o padrão é descrito por um conjunto de atributos. O algoritmo J48 foi executado utilizando-se a opção *cross-validation*, com o valor para *fold* igual a 10. Segundo Cordeiro, Hora e Mussa (2018), nessa opção, os dados são divididos em k conjuntos que se excluem entre si. Com k validações, são realizados k testes diferentes, enquanto o treinamento é realizado com k - 1 conjuntos remanescentes. Por fim, calcula-se a média das k validações. O número mínimo de objetos por folha da árvore de decisão foi ajustado para 5% do total de linhas da base de dados e como desfecho foi utilizado a divisão das médias em quartis (Q1, Q2, Q3 e Q4).

4.5. Resultados

Nesta seção serão apresentados os resultados obtidos após a execução da etapa de DM descrita na metodologia.

4.5.1. Países com alto desempenho em Matemática

A Tabela 4.2 demonstra as taxas de acertos obtidas com a aplicação do método de classificação J48. O algoritmo foi aplicado em cada base de dados dos países com alto desempenho na área de conhecimento Matemática no PISA 2015, de acordo com a OCDE (2016a). Para aplicação do método foi definido como desfecho a média em Matemática, obtida por cada país no PISA 2015, dividida em quartis.

Tabela 4.2 – Taxas de acerto para os países com alto desempenho em Matemática

Colocação no PISA 2015	Países	Taxa de acerto nos quartis (%)			
		Q1	Q2	Q3	Q4
1	Singapura	49,1	34,7	33,1	48,9
2	Hong Kong	50,7	31,8	29,6	43,1
3	Macau (China)	49,5	36,4	31,2	44,5
4	Taipei Chinesa	56,0	32,7	33,7	50,2
5	Japão	49,1	32,5	32,7	46,3

Fonte: Elaboração própria (2018).

O melhor resultado obtido foi para Taipei Chinesa, com 56% no quartil 1, e o pior resultado foi para Hong Kong (China), com 29,6 % no quartil 3. Os demais resultados variaram entre 31,2% e 50,7%. As árvores de decisão geradas são apresentadas nas Figuras 4.4, 4.5, 4.6, 4.7 e 4.8 do Apêndice A. No Quadro 4.1 é apresentado o atributo raiz e os demais atributos relevantes considerando a distância até o atributo raiz. Os dados estão relacionados com as folhas e ramificações obtidas nas árvores de decisão.

Quadro 4.1 – Atributos gerados nas árvores de decisão: Matemática.

Países	Atributo raiz	Distância do atributo raiz					
		1	2	3	4	5	6
Singapura	ST111Q01TA	ST121Q01NA	HOMEPOS	IC009Q08TA	ST121Q02NA	-	-
		HOMEPOS	ESCS	-	-	-	-
			IC008Q12TA	MMINS	IC008Q03TA	HISCED	-
Hong Kong	ST111Q01TA	IC007Q01TA	ST111Q01TA	ST076Q11NA	PA006Q02TA	-	-
		REPEAT	BSMJ	ST076Q07NA	IC008Q02TA	-	-
				ST076Q09NA	ST012Q09NA	-	-
Macau (China)	REPEAT	ST111Q01TA	ST111Q01TA	IC011Q01TA	MMINS	ST032Q01NA	-
			ANXTEST	ST013Q01TA	-	-	-
Taipei Chinesa	ST111Q01TA	ST076Q10NA	ST059Q02TA	ST111Q01TA	ST013Q01TA	-	-
		ISCEDD	ST013Q01TA	-	-	IC008Q12TA	IC002Q01NA
Japão	ST111Q01TA	IC013Q04NA	ST039Q01NA	ESCS	ST119Q01NA	-	-
		IC013Q04NA	MMINS	ST039Q01NA	MOTIVAT	-	-

Fonte: Elaboração própria (2018).

No Quadro 4.2 é apresentada a descrição dos atributos gerados nas árvores de decisão e apresentados no Quadro 4.1.

Quadro 4.2 – Descrição dos Atributos gerados nas árvores de decisão: Matemática.

Atributos	Descrição - Perguntas	Descrição - Alternativas
ANXTEST	Índice de ansiedade	-
BSMJ	Índice de posição ocupacional esperada dos estudantes	-
ESCS	Índice de <i>status</i> social e cultural econômico	-
HISCED	Índice do maior grau de escolaridade dos pais	-
HOMEPOS	Índice de posse de bens domésticos	-
IC002Q01NA	Idade que tinha quando usou pela primeira vez um dispositivo digital	(1) 6 anos ou menos (2) 7-9 anos, (3) 10-12 anos, (4) 13 anos ou mais, (5) Eu nunca havia utilizado um dispositivo eletrônico até hoje.
IC007Q01TA	Tempo que utiliza <i>internet</i> fora da escola	(1) Não utilizo nunca, (2) 1-30 minutos por dia, (3) 31-60 minutos por dia, (4) Entre 1 hora e 2 horas por dia, (5) Entre 2 horas e 4 horas por dia, (6) Entre 4 horas e 6 horas por dia, (7) Mais que 6 horas por dia.
IC008Q02TA	Frequência que utiliza dispositivos eletrônicos fora da escola para jogar jogos colaborativos <i>online</i> -	(1) Nunca ou quase nunca, (2) 1 ou 2 vezes no mês, (3) 1 ou 2 vezes na semana, (4) Quase todo dia, (5) Todo dia.
IC008Q03TA	Frequência que utiliza dispositivos eletrônicos fora da escola para acessar <i>email</i>	
IC008Q12TA	Frequência que utiliza dispositivos eletrônicos fora da escola para compartilhar conteúdos criados pelo próprio aluno (ex: música, poesia, vídeos, programas de computador)	
IC009Q08TA	Se a escola disponibiliza <i>pen drive</i> para utilizar na própria instituição	(1) Sim, e eu uso, (2) Sim, mas eu não uso, (3) Não.
IC011Q01TA	Frequência que utiliza dispositivos eletrônicos dentro da escola para conversar <i>online</i>	(1) Nunca ou quase nunca, (2) 1 ou 2 vezes no mês, (3) 1 ou 2 vezes na semana, (4) Quase todo dia, (5) Todo dia.
IC013Q04NA	Se o aluno concorda com a afirmação: “A <i>Internet</i> é um excelente recurso para obter informações nas quais estou interessado (a)”	(1) Discordo plenamente, (2) Discordo, (3) Concordo, (4) Concordo plenamente.
ISCEDD	Índice relacionado ao programa de estudo especial que o aluno frequentou e que visa dar acesso direto ao mercado de trabalho	(1) Educação acadêmica continuada, (2) Educação técnica avançada (3) Mercado de trabalho.
MMINS	Índice relacionado ao tempo de aprendizagem de Matemática (minutos por semana)	-

MOTIVAT	Índice de motivação estudantil	-
PA006Q02TA	Se a instituição ter uma boa reputação é uma razão para escolher uma escola para o filho	(1) Não é importante, (2) Um pouco importante, (3) Importante, (4) Muito importante.
REPEAT	Índice de reprovação	-
ST012Q09NA	Quantidade de instrumentos musicais (ex.: violão, piano) que o aluno possui em casa	(1) Nenhum, (2) Um, (2) Dois ou, (3) Três ou mais.
ST013Q01TA	Quantidade de livros que o aluno possui em casa	(1) 0 a 10 livros, (2) 11 a 25 livros, (3) 26 a 100 livros, (4) 101 a 200 livros, (5) 201 a 500 livros, (6) Mais de 500 livros.
ST032Q01NA	Quantidade de dias, durante uma semana, que faz atividades físicas moderadas num total de pelo menos 60 minutos por dia (ex.: caminhada, subir escadas, andar de bicicleta até a escola, pular cordas)	(1) 0 dia, (2) 1 dia, (3) 2 dias, (4) 3 dias, (5) 4 dias, (6) 5 dias, (7) 6 dias, (8) 7 dias.
ST039Q01NA	Durante os últimos 12 meses, frequência que o aluno teve a seguinte experiência: Os professores me chamaram com menos frequência com que eles chamaram outros alunos	(1) Nunca ou quase nunca, (2) Poucas vezes no ano, (3) Poucas vezes no mês, (4) Uma vez por semana ou mais
ST059Q02TA	Número de aulas de Matemática por semana	-
ST076Q07NA	Nos últimos dias, encontrou-se com amigos ou conversou com amigos pelo telefone antes de ir para a escola	(1) Sim (2) Não
ST076Q09NA	Nos últimos dias, fez o serviço de casa ou cuidou de outros membros da família antes de ir para a escola	
ST076Q10NA	Nos últimos dias, realizou trabalho remunerado antes de ir para a escola	
ST076Q11NA	Nos últimos dias, exercitou-se ou praticou esporte antes de ir para a escola	
ST078Q02NA	Nos últimos dias, estudou ou fez tarefa de casa após sair da escola	
ST111Q01TA	Nível de escolaridade que o aluno espera alcançar	(1) Ensino Fundamental completo (até o 9º ano), (2) Educação Profissional de Nível Técnico, (3) Ensino Médio completo (até o 3º ano), (4) Educação Profissional de nível tecnológico, (5) Graduação (Bacharelado, Licenciatura),

		(6) Pós-graduação (mestrado ou doutorado).
ST119Q01NA	Até que ponto você discorda ou concorda com a afirmação: “Eu quero notas altas na maioria ou em todas as minhas matérias”	(1) Discordo plenamente, (2) Discordo, (3) Concordo, (4) Concordo plenamente.
ST121Q01NA	Se o aluno discorda ou concorda com a afirmação: “Tiago desiste facilmente quando enfrenta um problema e frequentemente não está preparado para suas aulas. Tiago está motivado.”	
ST121Q02NA	Se o aluno discorda ou concorda com a afirmação: “Ana, na maior parte do tempo, permanece interessada nos trabalhos que começa e, às vezes, faz mais do que é esperado dela. Ana está motivada.”	

Fonte: Elaboração própria (2018).

As análises a seguir foram realizadas com base nas árvores de decisão apresentadas nas Figuras 4.4, 4.5, 4.6, 4.7 e 4.8 do Apêndice A, juntamente com o Quadro 4.2.

O atributo raiz ST111Q01TA foi semelhante em 4 dos 5 países com maior desempenho em Matemática, sendo diferente apenas em Macau/China. Em Macau, esse atributo ficou na distância 2 do atributo raiz e foi repetido na distância 1, conforme mostrado no Quadro 4.1. No PISA 2015, as questões com iniciais ST111 estão associadas ao nível de aspiração educacional dos alunos (OCDE, 2017a). Verificou-se que os alunos que aspiram alcançar a graduação e pós-graduação tenderam a alcançar os quartis com as médias mais altas em Matemática.

Em Macau, o atributo que apareceu como raiz foi o REPEAT (índice de reprovação). Tal atributo também foi encontrado na distância 1 em relação ao atributo raiz em Hong Kong/China, conforme mostrado no Quadro 4.1. Segundo OCDE (2017a), o atributo REPEAT é uma variável de repetência de ano que foi calculada pela recodificação das variáveis ST127Q01TA, ST127Q02TA e ST127Q03TA, que representam em qual nível o aluno já repetiu de ano (fundamental I, fundamental II ou médio). Esse atributo assume o valor de “1” se o aluno já repetiu de ano em pelo menos um nível e o valor de “0” se ele nunca repetiu. Constatou-se que, se o aluno já repetiu de ano em algum nível, o desfecho foi direto no quartil 1, no caso de Macau e no quartil 2, no caso de Hong Kong. Tais quartis representaram as médias mais baixas naquele país.

Em OCDE (2017a), verificou-se uma forte relação dos atributos ST012Q09NA (quantidade de instrumentos musicais), ST013Q01TA (quantidade de livros que o aluno possui em casa), HOMEPOS (índice de posse de bens domésticos), HISCED (índice do maior grau de escolaridade dos pais), BSMJ (índice de posição ocupacional esperada dos estudantes) e ESCS (índice de status

social e cultural econômico), todos apresentados no Quadro 4.1, com um total de 11 incidências. O atributo HOMEPOS é um índice resumido de todas as questões sobre itens domésticos e de posse (questões com iniciais ST011, ST012 e ST013). O HOMEPOS também é um dos três componentes na construção do índice de *status* econômico, social e cultural (ESCS). Para o cálculo do índice ESCS, também é utilizado o atributo HISCED. Segundo OCDE (2017a), como nos ciclos anteriores do PISA, os estudantes foram solicitados a relatar sua ocupação esperada aos 30 anos e uma descrição do trabalho. As respostas foram recodificadas gerando pontuações para o *status* ocupacional esperado dos alunos (BSMJ), onde valores mais altos indicaram níveis mais altos de *status* ocupacional esperado. Tal atributo também é um dos componentes utilizados para o cálculo do índice socioeconômico ESCS (OCDE, 2017a). Verificou-se que os valores mais baixos de todos os atributos mencionados tenderam a alcançar os quartis que representaram as médias mais baixas do desempenho em Matemática.

De acordo com OCDE (2017a), as questões ST076 e ST078 se referem às atividades que o aluno realizou antes e depois da escola. No Quadro 4.1 foi possível verificar que esses atributos tiveram 6 incidências nas árvores de decisão. A questão ST076 está relacionada às atividades que o aluno realizou antes de ir para a escola. Sobre a questão ST076, as respostas dos alunos que tiveram incidência foram:

- ST076Q07NA (encontrou-se com amigos ou conversou com amigos pelo telefone). Verificou-se que os alunos que responderam não a esse questionamento alcançaram os quartis que representam as médias mais altas.
- ST076Q09NA (fez o serviço de casa ou cuidou de outros membros da família) e ST076Q10NA (realizou trabalho remunerado). Verificou-se que os alunos que não realizam serviços em casa, bem como não exercem atividade remunerada, tenderam alcançar os quartis que representam as médias mais altas;
- ST076Q11NA (exercitou-se ou praticou esporte), que tem relação com o atributo ST032Q01NA (atividades físicas moderadas num total de pelo menos 60 minutos por dia), que teve uma incidência no Quadro 4.1. Verificou-se que os alunos que praticam atividade física tenderam alcançar os quartis que representam as médias mais altas.

No mesmo grupo acima, tem-se ainda a questão ST078 que está relacionada às atividades que o aluno realizou após sair da escola. A resposta do aluno que teve incidência foi a ST078Q02NA (estudou ou fez tarefa de casa). Verificou-se que os alunos que não fazem tarefa de casa alcançaram as médias representadas nos quartis 2 e 3.

Segundo a OCDE (2017a), o atributo MMINS (tempo de aprendizagem em Matemática), que teve três incidências no Quadro 4.1, está associado ao atributo ST059Q02TA (número de aulas de Matemática por semana), que teve apenas 1 incidência. Verificou-se que os alunos que alcançaram as maiores médias tiveram um tempo maior de aula de Matemática durante a semana.

Os atributos ST119Q01NA, ST121Q01NA, ST121Q02NA e MOTIVAT são relacionados à motivação estudantil e o atributo ANXTEST é relacionado a ansiedade. Juntos, esses atributos tiveram 5 incidências no Quadro 4.1. De acordo com a OCDE(2016b), o envolvimento dos alunos com a aprendizagem e a motivação são fatores que estão associados ao desempenho do aluno e às atitudes em relação ao processo da aprendizagem. Verificou-se que os alunos motivados e com baixo de índice de ansiedade tenderam a alcançar os quartis que representam as médias mais altas.

Os atributos IC002Q01NA, IC007Q01TA, IC008Q02TA, IC008Q03TA, IC008Q12TA, IC009Q08TA, IC011Q01TA e IC013Q04NA apresentados no Quadro 4.1, são relacionados às tecnologias de informação e comunicação (TIC's). Verificou-se que o aluno que nunca utilizou *internet* ou que a utiliza menos de 30 minutos por dia fora da escola (IC007Q01TA) juntamente com aquele aluno que utiliza *internet* para conversar *online* com os colegas na escola (IC011Q01TA) atingiram os quartis que representam as menores médias do desempenho em Matemática. Por outro lado, os alunos que começaram a utilizar os dispositivos digitais com idade inferior a 10 anos (IC002Q01NA) juntamente com aqueles alunos que afirmaram que a *internet* é um excelente recurso para obter informações (IC013Q04NA) e que utilizam dispositivos digitais para jogos colaborativos (IC008Q02TA), para compartilhar dados (IC008Q12TA) e para acessar *e-mail* (IC008Q03TA), tenderam a alcançar os quartis que representam as médias mais altas.

Os atributos PA006Q02TA, ISCEDD e ST039Q01NA tiveram apenas 1 incidência no Quadro 4.1. O atributo PA006Q02TA está relacionado ao fato dos pais considerarem que a instituição ter uma boa reputação é uma razão para escolher a escola para o filho. Verificou-se que se os pais afirmaram que é muito importante, os filhos alcançaram o quartil 2, enquanto os demais alcançaram o quartil 1. O atributo ISCEDD indica o fluxo do programa de estudo do aluno, sendo as opções: educação acadêmica continuada, educação técnica avançada ou mercado de trabalho. Constatou-se que esse atributo foi relevante para se alcançar as médias representadas nos quartis 2, 3 e 4. O atributo ST039Q01NA (frequência que o aluno teve a seguinte experiência: “Os professores me chamaram com menos frequência com que eles chamaram outros alunos”) está relacionado a imparcialidade do professor. Verificou-se que esse atributo foi relevante para se alcançar as médias representadas nos quartis 3 e 4.

4.5.2. Países com alto desempenho em Leitura

A Tabela 4.3 demonstra as taxas de acertos obtidas com a aplicação do método de classificação J48. O algoritmo foi aplicado em cada base de dados dos 5 países com maior desempenho na área de conhecimento Leitura no PISA 2015, de acordo com a OCDE (2016a). Para aplicação do método foi definido como desfecho a média em Leitura, obtida por cada país no PISA 2015, dividida em quartis.

Tabela 4.3 – Taxas de acerto para os países com alto desempenho em Leitura

Colocação no PISA 2015	Países	Taxa de acerto nos quartis (%)			
		Q1	Q2	Q3	Q4
1	Singapura	48,9	31,4	30,5	45,8
2	Hong Kong	51,3	34,8	35,0	46,3
3	Canadá	47,0	31,5	32,0	47,0
4	Finlândia	47,7	30,6	28,2	45,0
5	Irlanda	49,9	33,7	32,2	51,3

Fonte: Elaboração própria (2018).

O melhor resultado obtido foi para Hong Kong, no quartil 1, e Irlanda, no quartil 4, ambos com 51,3%, e o pior resultado foi para Finlândia com 28,2% no quartil 3. Os demais resultados variaram entre 30,5% e 49,9%. As árvores de decisão geradas são apresentadas nas Figuras 4.9, 4.10, 4.11, 4.12 e 4.13 do Apêndice A. No Quadro 4.3 é apresentado o atributo raiz e os demais atributos relevantes considerando a distância até o atributo raiz. Os dados estão relacionados com as folhas e ramificações obtidas nas árvores de decisão com desfecho na média em Leitura.

Quadro 4.3 – Atributos gerados nas árvores de decisão: Leitura.

Países	Atributo raiz	Distância do atributo raiz						
		1	2	3	4	5	6	7
Singapura	ST111Q01TA	ST121Q01NA	HOMEPOS	ST076Q11NA	LMINS	-	-	-
		HOMEPOS	S0IAICT	-	-	-	-	-
Hong Kong	ST121Q01NA	ST111Q01TA	ST076Q11NA	ST111Q01TA	ST111Q01TA	-	-	-
				ST001D01T	S0IAICT	-	-	-
				HADDINST	-	-	-	-
Canadá	ST111Q01TA	ST121Q03NA	ST013Q01TA	ST076Q11NA	-	-	-	
		ST076Q10NA	ESCS	-	-	-	-	
Finlândia	ST111Q01TA	ST121Q02NA	IC013Q04NA	USESCH	ST013Q01TA	ST013Q01TA	ST121Q03NA	ST118Q02NA
						ST082Q13NA	-	-
Irlanda	ST013Q01TA	ST111Q01TA	ST078Q10NA	ST121Q01NA	ST013Q01TA	-	-	-
			ST082Q13NA	ST121Q02NA	ST013Q01TA	-	-	-

Fonte: Elaboração própria (2018).

No Quadro 4.4 é apresentada a descrição dos atributos gerados nas árvores de decisão e já apresentados no Quadro 4.3.

Quadro 4.4 – Descrição dos Atributos gerados nas árvores de decisão: Leitura.

Atributos	Descrição - Perguntas	Descrição - Alternativas
CPSVALUE	Índice relacionado a colaboração e trabalho em equipe: cooperação de valor	-
ESCS	Índice de status social e cultural econômico	-
HADDINST	Índice relacionado as horas de estudos adicionais	-
HOMEPOS	Índice de posse de bens domésticos	-
IC008Q03TA	Frequência que utiliza dispositivos eletrônicos fora da escola para acessar <i>email</i>	(1) Nunca ou quase nunca, (2) 1 ou 2 vezes no mês, (3) 1 ou 2 vezes na semana, (4) Quase todo dia, (5) Todo dia
IC013Q04NA	Se o aluno concorda com a afirmação: “A <i>Internet</i> é um excelente recurso para obter informações nas quais estou interessado (a)”	(1) Discordo plenamente, (2) Discordo, (3) Concordo, (4) Concordo plenamente
LMINS	Índice relacionado ao tempo de aprendizagem de Leitura (minutos por semana)	-
SOLIACT	Índice relacionado ao grau em que as Tecnologias de Informação e Comunicação (TIC) fazem parte de sua vida social diária	-
ST001D01T	Série/ano que o aluno se encontra	-
ST013Q01TA	Quantidade de livros que o aluno possui em casa	(1) 0 a 10 livros, (2) 11 a 25 livros, (3) 26 a 100 livros, (4) 101 a 200 livros, (5) 201 a 500 livros, (6) Mais de 500 livros
ST076Q10NA	Nos últimos dias, realizou trabalho remunerado antes de ir para a escola	(1) Sim (2) Não
ST076Q11NA	Nos últimos dias, exercitou-se ou praticou esporte antes de ir para a escola	
ST078Q10NA	Nos últimos dias, realizou trabalho remunerado após sair da escola	
ST082Q13NA	Até que ponto o aluno discorda ou concorda com a afirmação: “Eu acho que trabalho em equipe aumenta minha própria eficiência”	(1) Discordo plenamente, (2) Discordo, (3) Concordo (4) Concordo plenamente
ST111Q01TA	Nível de escolaridade que o aluno espera alcançar	(1) Ensino Fundamental completo (até o 9º ano), (2) Educação Profissional de Nível Técnico, (3) Ensino Médio completo (até o 3º ano), (4) Educação Profissional de nível tecnológico,

		(5) Graduação (Bacharelado, Licenciatura), (6) Pós-graduação (mestrado ou doutorado).
ST118Q02NA	Até que ponto o aluno discorda ou concorda com a afirmação: “Eu me preocupo se tirarei notas baixas na escola”	(1) Discordo plenamente, (2) Discordo, (3) Concordo, (4) Concordo plenamente.
ST121Q01NA	Se o aluno discorda ou concorda com a afirmação: “Tiago desiste facilmente quando enfrenta um problema e frequentemente não está preparado para suas aulas. Tiago está motivado.”	
ST121Q02NA	Se o aluno discorda ou concorda com a afirmação: “Ana, na maior parte do tempo, permanece interessada nos trabalhos que começa e, às vezes, faz mais do que é esperado dela. Ana está motivada.”	
ST121Q03NA	Se o aluno discorda ou concorda com a afirmação: “Lúcia quer notas altas na escola e continua melhorando os trabalhos até que tudo esteja perfeito. Lúcia está motivada.”	
USESCH	Índice relacionado ao uso de TIC na escola	-

Fonte: Elaboração própria (2018).

As análises a seguir foram realizadas com base nas árvores de decisão apresentadas nas Figuras 4.9, 4.10, 4.11, 4.12 e 4.13 do Apêndice A, juntamente com o Quadro 4.4.

Verificou-se no Quadro 4.3 que o atributo raiz ST111Q01TA, que representa a aspiração do nível educacional dos alunos, foi semelhante em 3 dos 5 países com maior desempenho em Leitura no PISA 2015, sendo diferente apenas em Hong Kong e Irlanda. Nesses países, o atributo mencionado se apresenta na distância 1 do atributo raiz. Tal atributo também teve incidência no Quadro 4.1 apresentado na Seção 4.5.1, que trata dos países com maior desempenho em Matemática. Constatou-se que os alunos que esperam alcançar o nível da graduação ou pós-graduação tenderam a alcançar os quartis que representam as médias mais altas em Leitura.

Em Hong Kong, o atributo que se apresenta como raiz é o ST121Q01NA, que junto com os atributos ST121Q02NA e ST121Q03NA, são relacionados a motivação estudantil e tiveram 8 incidências no total no Quadro 4.3. Tal atributo tem conexão com o atributo ST118Q02NA, que teve 1 incidência e relaciona-se com a ansiedade. A questão da ansiedade e motivação foi mencionada também na Seção 4.5.1, que trata dos países com maior desempenho em Matemática. Verificou-se que alunos desmotivados tenderam a alcançar os quartis que representam as menores médias naquele país.

Na Irlanda, o atributo que se apresenta como raiz é o ST013Q01TA (quantidade de livros que o aluno possui em casa), que teve mais 6 incidências no Quadro 4.3. Conforme mencionado na Seção 4.5.1, esse atributo está relacionado ao atributo HOMEPOS (índice de posse de bens domésticos), que teve 2 incidências no Quadro 4.3. O atributo HOMEPOS é um dos componentes para o cálculo do índice de *status* socioeconômico ESCS. Constatou-se que os alunos que tem um número elevado de livros em casa e aqueles que apresentam o índice de posses domésticas alto, tenderam a alcançar os quartis que representam as maiores médias em Leitura. Tais atributos também tiveram incidência no Quadro 4.1, apresentado na Seção 4.5.1, que trata dos países com maior desempenho em Matemática.

Os atributos IC008Q03TA, IC013Q04NA, SOIAICT e USESCH apresentados no Quadro 4.3 são relacionados ao uso das TIC's dentro e fora da escola. Constatou-se que os alunos que utilizam a *internet* para acessar *email* fora da escola (IC008Q03TA) e também aqueles que concordam que a *internet* é um excelente recurso para obter informações nas quais está interessado (IC013Q04NA) tenderam a alcançar os quartis com as médias mais altas. Por outro lado, verificou-se que aqueles alunos que têm um alto índice de uso das tecnologias de informação e comunicação (SOIAICT e USESCH) tenderam a alcançar os quartis das menores médias. Os atributos relacionados às TIC's também tiveram incidência no Quadro 4.1, apresentado na Seção 4.5.1, que trata dos países com maior desempenho em Matemática.

De acordo com a OCDE (2017a) o atributo CPSVALUE, descrito como um índice relacionado a colaboração e trabalho em equipe, foi calculado utilizando as questões com iniciais ST082. Assim, o atributo ST082Q13NA está relacionado ao atributo CPSVALUE. Ambos atributos tiveram 3 incidências no Quadro 4.3. Verificou-se que a cooperação e o trabalho em equipe (CPSVALUE) atuaram de forma positiva levando os alunos a alcançarem os quartis que representam as médias mais altas. Por outro lado, verificou-se que os alunos que alcançaram o quartil das médias mais altas não acreditam que o trabalho em equipe aumenta a própria eficiência (ST082Q13NA). Realizando uma análise mais profunda diretamente na base de dados, constatou-se que esses alunos concordam que é melhor tomar decisões em equipes do que individualmente (ST082Q09NA) e afirmaram que gostam de trabalhar em equipe, colaborando com os colegas de classe (ST082Q14NA) e de vê-los bem-sucedidos (ST082Q03NA).

Os atributos ST076Q10NA e ST078Q10NA, apresentados no Quadro 4.3 referem-se à realização de trabalho remunerado antes e depois da escola. Verificou-se que o exercício de atividade remunerada atuou de forma negativa no desempenho dos alunos. O atributo ST076Q11NA está

relacionado a prática de esportes antes de ir para a escola. Tal atributo teve 5 incidências no Quadro 4.3. Constatou-se que os alunos que não praticam esportes antes de ir para a escola tenderam a alcançar os quartis das médias mais altas. Por outro, realizando uma análise mais profunda diretamente na base de dados, verificou-se que grande parte desses alunos afirmaram que realizam atividade física após sair da escola (ST076Q11NA) e também que praticam atividades físicas moderadas num total de pelo menos 60 minutos por dia por mais de 2 dias na semana (ST032Q01NA).

Os atributos ST001D01T, LMINS e HADDINST tiveram apenas uma incidência no Quadro 4.3. Sobre os atributos ST001D01T, que está relacionado a série/ano que o aluno se encontra, e LMINS, que está relacionado ao tempo de aprendizagem de Leitura (minutos por semana), verificou-se que esses atributos foram relevantes para se alcançar os quartis 2 e 3. E quanto ao atributo HADDINST, que está relacionado as horas de estudos adicionais, constatou-se que o fato do aluno assistir às aulas adicionais, além das aulas obrigatórias, foi relevante para se alcançar os quartis 3 e 4.

4.5.3. Países com alto desempenho em Ciências

A Tabela 4.4 demonstra as taxas de acertos obtidas com a aplicação do método de classificação J48. O algoritmo foi aplicado em cada base de dados dos 5 países com maior desempenho na área de conhecimento Ciências no PISA 2015, de acordo com a OCDE (2016a). Para aplicação do método foi definido como desfecho a média em Ciências, obtida por cada país no PISA 2015, dividida em quartis.

Tabela 4.4 – Taxas de acerto para os países com alto desempenho em Ciências

Colocação no PISA 2015	Países	Taxa de acerto nos quartis (%)			
		Q1	Q2	Q3	Q4
1	Singapura	60,9	38,8	31,1	51,5
2	Japão	49,1	32,8	34,4	51,3
3	Estônia	53,7	35,8	33,4	54,9
4	Taipei Chinesa	56,2	35,6	35,8	58,4
5	Finlândia	53,6	34,3	34,2	50,4

Fonte: Elaboração própria (2018).

O melhor resultado obtido foi para Singapura, com 60,9% no quartil 1, e o pior resultado foi para também para Singapura, com 31,1% no quartil 3. Os demais resultados variaram entre 32,8% e 58,4%. As árvores de decisão geradas são apresentadas nas Figuras 4.14, 4.15, 4.16, 4.17 e 4.18 do

Apêndice A. No Quadro 4.5 é apresentado o atributo raiz e os demais atributos relevantes considerando a distância até o atributo raiz. Os dados estão relacionados com as folhas e ramificações obtidas nas árvores de decisão com desfecho na média em Ciências.

Quadro 4.5 – Atributos gerados nas árvores de decisão: Ciências.

Países	Atributo raiz	Distância do atributo raiz				
		1	2	3	4	5
Singapura	ST092Q01TA	ST076Q10NA	SMINS	HOMEPOS	ST093Q01TA	-
				ST063Q02NA	ST131Q06NA	IC001Q08TA
Japão	ST092Q01TA	ST111Q01TA	ST092Q05TA	IC010Q01TA	-	-
		ST111Q01TA	EPIST	ST078Q11NA	-	-
Estônia	ST111Q01TA	ST121Q01NA	ST092Q01TA	ST076Q11NA	-	-
		ST092Q01TA	ST076Q11NA	-	-	-
Taipei Chinesa	ST092Q01TA	ST092Q05TA	ST013Q01TA	-	-	-
		ST111Q01TA	ST131Q04NA	ST111Q01TA	-	-
Finlândia	ST092Q01TA	ST121Q03NA	ST118Q02NA	-	-	-
		ST111Q01TA	ST129Q02TA	UNFAIRTEACHER	SMINS	-
			EPIST	HOMESCH	SMINS	-
				ST129Q02TA	-	-
				ISCEDD	EPIST	IC008Q11TA
				ESCS	ESCS	ESCS

Fonte: Elaboração própria (2018).

No Quadro 4.6 é apresentada a descrição dos atributos gerados nas árvores de decisão e já apresentados no Quadro 4.5.

Quadro 4.6 – Descrição dos Atributos gerados nas árvores de decisão: Ciências.

Atributos	Descrição - Perguntas	Descrição - Alternativas
EPIST	Índice relacionado a opiniões dos estudantes sobre as abordagens científicas	-
ESCS	Índice de status social e cultural econômico	-
HOMEPOS	Índice de posse de bens domésticos	-
HOMESCH	Índice relacionado ao uso de TIC fora da escola para trabalhos escolares	-
IC001Q08TA	Se o aluno tem leitor de música portátil (Mp3/Mp4 player, iPod ou similar) disponível para utilizar em casa	(1) Sim, e eu uso, (2) Sim, mas eu não uso, (3) Não.
IC008Q11TA	Frequência que utiliza dispositivos eletrônicos fora da escola para baixar músicas, filmes, jogos ou programas pela <i>Internet</i> .	(1) Nunca ou quase nunca, (2) 1 ou 2 vezes no mês, (3) 1 ou 2 vezes na semana,
IC010Q01TA	Frequência que utiliza dispositivos eletrônicos fora da escola para navegar na <i>Internet</i> para fazer trabalhos escolares (ex: para preparar uma apresentação)	(4) Quase todo dia, (5) Todo dia

IC013Q04NA	Se o aluno concorda com a afirmação: “A <i>Internet</i> é um excelente recurso para obter informações nas quais estou interessado (a)”	(1) Discordo plenamente, (2) Discordo, (3) Concordo, (4) Concordo plenamente
ISCEDD	Índice relacionado ao programa de estudo especial que o aluno frequentou e que visa dar acesso direto ao mercado de trabalho	(1) Educação acadêmica continuada, (2) Educação técnica avançada (3) Mercado de trabalho.
JOYSCIE	Índice de apreciação de ciência	-
SMINS	Índice relacionado ao tempo de aprendizagem em Ciências (minutos por semana)	-
ST013Q01TA	Quantidade de livros que o aluno possui em casa	(1) 0 a 10 livros, (2) 11 a 25 livros, (3) 26 a 100 livros, (4) 101 a 200 livros, (5) 201 a 500 livros, (6) Mais de 500 livros
ST063Q02NA	Se o aluno teve aula de química	(1) Este ano, (2) No ano passado
ST076Q10NA	Nos últimos dias, realizou trabalho remunerado antes de ir para a escola	(1) Sim (2) Não
ST076Q11NA	Nos últimos dias, exercitou-se ou praticou esporte antes de ir para a escola	
ST078Q11NA	Nos últimos dias, exercitou-se ou praticou esporte após sair da escola	
ST092Q01TA	Até que ponto o aluno está informado em relação a questão ambiental: A redução da camada de ozônio na atmosfera	(1) Nunca ouvi nada a respeito disso, (2) Já ouvi algo sobre isso, mas não seria capaz de explicar o que realmente significa, (3) Sei alguma coisa sobre isso e poderia explicar em termos gerais, (4) Estou familiarizado com o assunto e seria capaz de explicar perfeitamente
ST092Q05TA	Até que ponto o aluno está informado em relação a questão ambiental: Devastação de florestas para outros usos da terra	
ST093Q01TA	Se o aluno acha que os problemas associados à poluição do ar vão melhorar ou piorar nos próximos 20 anos	(1) Melhorar, (2) Permanecer inalterados, (3) Piorar.
ST111Q01TA	Nível de escolaridade que o aluno espera alcançar	(1) Ensino Fundamental completo (até o 9º ano), (2) Educação Profissional de Nível Técnico, (3) Ensino Médio completo (até o 3º ano), (4) Educação Profissional de nível tecnológico, (5) Graduação (Bacharelado, Licenciatura), (6) Pós-graduação (mestrado ou doutorado).
ST118Q02NA	Até que ponto o aluno discorda ou concorda com a afirmação: “Eu me preocupo se tirarei notas baixas na escola”	(1) Discordo plenamente, (2) Discordo, (3) Concordo,

ST121Q01NA	Se o aluno discorda ou concorda com a afirmação: “Tiago desiste facilmente quando enfrenta um problema e frequentemente não está preparado para suas aulas. Tiago está motivado.”	(4) Concordo plenamente.
ST121Q03NA	Se o aluno discorda ou concorda com a afirmação: “Lúcia quer notas altas na escola e continua melhorando os trabalhos até que tudo esteja perfeito. Lúcia está motivada.”	
ST129Q02TA	Facilidade que o aluno realizaria a seguinte tarefa sozinho(a): Explicar por que os terremotos ocorrem com mais frequência em algumas áreas e não em outras	(1) Eu conseguiria realizar facilmente, (2) Eu conseguiria realizar com um pouco de esforço, (3) Eu teria que me esforçar muito para realizar isto sozinho(a), (4) Eu não conseguiria realizar
ST131Q04NA	Se o aluno discorda ou concorda com a afirmação: “Boas respostas são baseadas em provas de diferentes experimentos.”	(1) Discordo plenamente, (2) Discordo, (3) Concordo, (4) Concordo plenamente.
ST131Q06NA	Se o aluno discorda ou concorda com a afirmação: “É bom fazer experiências mais de uma vez para certificar as suas descobertas.”	
UNFAIRTEACHER	Índice relacionado a imparcialidade do professor	-

Fonte: Elaboração própria (2018).

As análises a seguir foram realizadas com base nas árvores de decisão apresentadas nas Figuras 4.14, 4.15, 4.16, 4.17 e 4.18 do Apêndice A, juntamente com o Quadro 4.6.

O atributo raiz ST092Q01TA foi semelhante em 3 dos 5 países com melhor desempenho em Ciências no PISA 2015, sendo diferente apenas na Estônia. No total esse atributo teve 8 incidências no Quadro 4.5, e na Estônia aparece na distância 1 do atributo raiz. Tal atributo está relacionado ao nível de informação que o aluno tem sobre a questão ambiental: Redução da camada de ozônio na atmosfera. Verificou-se que se os alunos sabem algo sobre o assunto e podem explicar em termos gerais, ou se os alunos estão familiarizados com o tema e seriam capazes de explicar perfeitamente foi relevante para se alcançar os quartis que representam as médias mais altas em Ciências. O atributo ST092Q01TA tem conexão com o atributo ST093Q01TA, que está relacionado a opinião do aluno sobre os problemas ambientais associados à poluição do ar. Verificou-se que aquele aluno que acredita que a situação vai piorar nos próximos 20 anos alcançou média mais alta do que aquele aluno que acredita que a situação vai melhorar ou vai permanecer inalterada nos próximos 20 anos.

Na Estônia, o atributo que se apresentou como raiz foi o ST111Q01TA, que representa a aspiração do aluno sobre o nível educacional. No total esse atributo teve 7 incidências no Quadro 4.5.

Constatou-se que os alunos que aspiram alcançar a graduação ou pós-graduação tenderam a alcançar os quartis que representam as médias mais altas em Ciências. Tal atributo também teve incidência na análise dos países com maior desempenho em Matemática e Leitura, apresentada na Seções 4.5.1 e 4.5.2.

O atributo SMINS que está relacionado o número de aulas de Ciências por semana tem conexão com o atributo ST063Q02NA que está relacionado se o aluno teve aula de química. Por se tratar da área de conhecimento de Ciências, esses atributos também têm conexão com os atributos JOYSCIE e ST129Q02TA. O atributo JOYSCIE está relacionado com as questões com iniciais ST094, que versam sobre se o aluno se diverte, gosta de ler, fica feliz, gosta de adquirir novos conhecimentos sobre Ciências e tem interesse sobre essa área. Esses atributos tiveram 5 incidências no Quadro 4.5. Verificou-se que os alunos que alcançaram as médias representadas nos quartis mais altos foram aqueles que tem mais aulas de Ciências durante a semana, que têm um alto índice de apreciação por essa área de conhecimento e que sabem explicar facilmente a frequência de ocorrência de terremotos em algumas áreas.

Ainda no contexto das abordagens específicas da área da ciência, verificou-se a incidência do atributo EPIST (índice de confiança sobre abordagens científicas), que de acordo com a OCDE (2017a), está relacionado aos atributos ST131Q04NA (se o aluno concorda com a afirmação: “Boas respostas são baseadas em provas de diferentes experimentos.”) e ST131Q06NA (se o aluno concorda com a afirmação: “É bom fazer experiências mais de uma vez para certificar as suas descobertas.”). Esses atributos tiveram 5 incidências no Quadro 4.5 e tratam da confiança epistemológica do aluno sobre o tema. Averiguou-se que se o aluno concorda que boas respostas são baseadas em provas de diferentes experimentos (ST131Q04NA), se ele acredita que é bom fazer experiências mais de uma vez para certificar as suas descobertas (ST131Q06NA) e o alto índice de confiança sobre abordagens científicas (EPIST) foram relevantes para se alcançar os quartis das maiores médias em Ciências.

As questões ST076 e ST078 se referem às atividades que o aluno realizou antes e depois da escola. No total, essas questões tiveram 5 incidências no Quadro 4.5. Em relação ao atributo ST076Q10NA que está relacionado a realização de trabalho remunerado, verificou-se que os alunos que não exercem atividade remunerada antes de ir para a escola, tenderam alcançar os quartis que representam as médias mais altas. Quanto aos atributos ST076Q11NA e ST078Q11NA, verificou-se que os alunos que responderam que não praticam atividade física antes ou depois da escola alcançaram médias mais altas do que aqueles que praticam. Por outro lado, realizando uma análise

mais profunda diretamente na base de dados constatou-se que esses alunos informaram que praticam atividade física por no mínimo 2 dias na semana, sendo 60 minutos por dia (ST032Q01NA).

Os atributos ST121Q01NA e ST121Q03NA são relacionados a motivação estudantil e têm conexão com o atributo ST118Q02NA. No total esses atributos tiveram 3 incidências no Quadro 4.5. Verificou-se que esses atributos tiveram impacto no caminho para se alcançar os quartis que representam as menores médias em Ciências. Os atributos relacionados a motivação estudantil também foram encontrados nas Seções 4.5.1 e 4.5.2, que tratam dos países com maior desempenho em Matemática e Leitura.

Segundo a OCDE (2017a) os atributos HOMESCH, IC001Q08TA, IC008Q11TA, IC010Q01TA e IC013Q04NA são relacionados ao uso TIC's. No total esses atributos tiveram 5 incidências no Quadro 4.5. Verificou-se que o fato do aluno concordar que a *internet* é um excelente recurso para obter informações de que está interessado (IC013Q04NA) juntamente com o fato do aluno utilizar dispositivos eletrônicos fora da escola de forma consciente, seja leitor de música portátil (IC001Q08TA), seja para baixar músicas, filmes, jogos ou programas pela *Internet* (IC008Q11TA), seja para compartilhar conteúdos criados por ele (IC013Q04NA) ou para navegar na *Internet* para fazer trabalhos escolares (IC010Q01TA) foi relevante para se alcançar os quartis que representam as maiores médias em Ciências. Os atributos relacionados ao uso das tecnologias de informação e comunicação também foram encontrados nas Seções 4.5.1 e 4.5.2, que tratam dos países com maior desempenho em Matemática e Leitura.

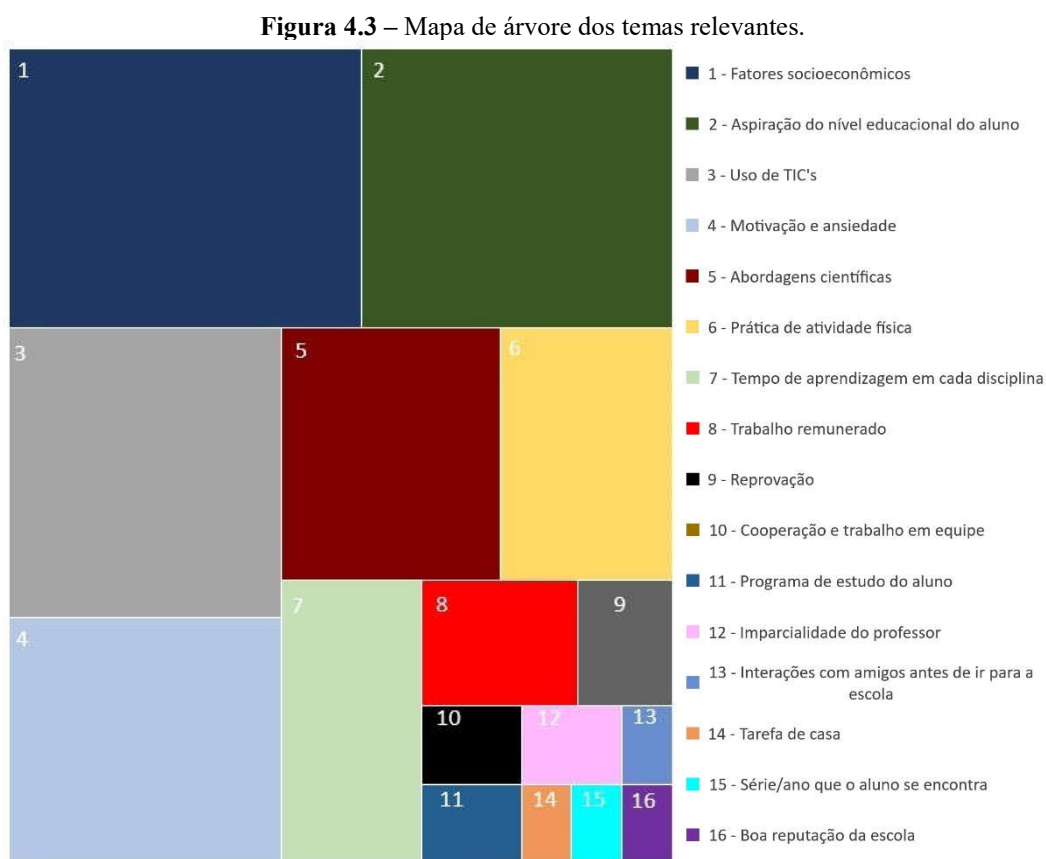
Os atributos HOMEPOS (índice de posse de bens domésticos) e ST013Q01TA (quantidade de livros que o aluno possui em casa) são relacionados ao índice socioeconômico ESCS. Juntos, esses atributos tiveram 4 incidências no Quadro 4.5. Constatou-se que os alunos que não tiveram um bom desempenho na área de Ciências foram aqueles que têm menos de 25 livros em casa. Por outro lado, os alunos que alcançaram as maiores médias em Ciências foram aqueles que tem um alto índice de posse de bens domésticos e socioeconômico. Esses atributos também foram encontrados nas árvores de decisão dos países com maior desempenho em Matemática e Leitura, apresentadas nas Seções 4.5.1 e 4.5.2.

Os atributos ISCEDD (Índice relacionado ao programa de estudo especial) e UNFAIRTEACHER (Índice relacionado a imparcialidade do professor) tiveram apenas 1 incidência no Quadro 4.5. Verificou-se que esses atributos foram relevantes para se alcançar os quartis 3 e 4, que representam as maiores médias em Ciências. O atributo UNFAIRTEACHER tem relação com o atributo ST039Q01NA (frequência que o aluno teve a seguinte experiência: “Os professores me

chamaram com menos frequência com que eles chamaram outros alunos”), que junto com o atributo ISCEDD também podem ser encontrados na Seção 4.5.1, que trata dos países com maior desempenho em Matemática.

4.6. Análise e Discussão de Resultados

A partir dos resultados encontrados, foi realizada a análise dos atributos que se apresentaram como raiz, dos atributos que tiveram incidência em mais de 1 área de conhecimento, dos atributos que são relacionados entre si e por último dos atributos que tiveram incidência somente em 1 área de conhecimento. A Figura 4.3 apresenta um mapa de árvore, que é uma visualização mais estruturada dos temas que foram relevantes nos países que se destacaram no PISA 2015. No mapa de árvore apresentado é possível verificar, pelo tamanho dos retângulos, a predominância do tema.



Fonte: Elaboração própria (2018).

Os atributos relacionados aos fatores socioeconômicos foram encontrados nas 3 áreas de conhecimento (Matemática, Leitura e Ciências) e apresentaram uma forte influência no desempenho

dos alunos, conforme mostra a Figura 4.3. No total, esses atributos obtiveram 25 incidências, sendo 1 apresentado como atributo raiz. O atributo ST013Q01TA aparece como raiz na área de Leitura, apresentada na Seção 4.5.2. Esse atributo juntamente com o atributo ST012Q09NA são relacionados ao atributo HOMEPOS, que é um dos componentes para o cálculo do índice de *status* socioeconômico ESCS. De acordo com a OCDE (2017a), para o cálculo do índice ESCS, também são utilizados os atributos HISCED e BSMJ. Verificou-se que os alunos que apresentaram um valor alto dos atributos relacionados aos fatores socioeconômicos tenderam a alcançar os quartis que representam as médias mais altas. Para Gonçalves (2017), os recursos sociais, incluindo a formação escolar dos pais e o nível socioeconômico, são relacionados com as oportunidades de aprendizagem e predizem o desempenho escolar dos alunos. Nesse sentido, Alves e Silva Filho (2017) afirmaram que em relação ao índice ESCS, as pesquisas indicaram que, quanto maior o nível socioeconômico, maior será o desempenho do aluno. Segundo a OCDE (2016c), nos países em que houve a implantação de políticas universais, envolvendo os alunos e escolas desfavorecidas se obteve uma melhoria no desempenho estudantil.

Considerando as 3 áreas de conhecimento, verificou-se que o atributo ST111Q01TA, descrito como a aspiração nível educacional do aluno, teve também uma forte relevância no desempenho dos alunos, conforme mostra a Figura 4.3. No total obteve 22 incidências e se apresentou como raiz em 8 das 15 árvores de decisão apresentadas. Constatou-se que os alunos que aspiram alcançar o nível da graduação ou pós-graduação tenderam a alcançar os quartis que representam as médias mais altas. Nesse sentido, Borgonov e Pál (2016) afirmaram que a necessidade de realização pessoal pode influenciar no desempenho acadêmico, escolha de carreira e desempenho no trabalho. Pessoas com altos níveis de motivação são mais propensas a escolher ocupações que permitam mais controle sobre os resultados. A motivação de conquistar o que se deseja desperta nos alunos o interesse de superar os outros e de serem bem-sucedidos em termos de competição com algum padrão de excelência.

Os atributos relacionados ao uso das tecnologias de informação e comunicação tiveram 20 incidências nas 3 áreas de conhecimento. Destacaram-se os atributos IC001Q08TA, IC002Q01NA, IC007Q01TA, IC008Q02TA, IC008Q03TA, IC008Q12TA, IC009Q08TA, IC008Q11TA, IC010Q01TA, IC011Q01TA, IC013Q04NA, HOMESCH, SOIAICT e USESCH. Sobre as tecnologias de informação e comunicação, constatou-se que os alunos que afirmaram que a *internet* é um excelente recurso para obter informações e que utilizam os dispositivos eletrônicos e a *internet* de forma adequada tenderam a alcançar os quartis que representam as médias mais altas. Por outro lado, aqueles alunos que têm um alto índice de uso das tecnologias de informação e comunicação tenderam a alcançar os quartis das menores médias. De acordo com a OCDE (2017b) os usuários

extremos da *internet* pontuaram cerca de 30 pontos a menos em todas as disciplinas. Dorn *et al.* (2017) ressaltaram que análises externas evidenciaram que acessar a *internet* com objetivos educacionais e para aprendizagem colaborativa, por meio de jogos, tem efeitos positivos, ao passo que a participação em mídias sociais parece ser negativa não apenas em termos das notas, mas também para o bem-estar do aluno. Para a OCDE (2017b) as escolas podem considerar investir em um plano abrangente de educação e supervisão para ajudar os alunos a adquirir os conhecimentos, habilidades e motivação que eles precisam para utilizar a *internet* com segurança e responsabilidade.

Os atributos ST119Q01NA, ST121Q01NA, ST121Q02NA, ST121Q03NA e MOTIVAT são relacionados a motivação estudantil. Desses, o atributo ST121Q01NA se apresentou como raiz na área de Leitura, apresentada na Seção 4.5.2. Esses atributos, têm conexão com os atributos ST118Q02NA e ANXTEST, que são relacionados a ansiedade. Juntos, esses atributos tiveram 17 incidências levando em consideração as 3 áreas de conhecimento. Verificou-se que alunos com alto índice de motivação e baixo índice de ansiedade tenderam a alcançar os quartis que representam as maiores médias. Segundo Borgonov e Pál (2016) a autoconfiança tem um impacto na aprendizagem e no desempenho cognitivo, motivacional, afetivo e na tomada de decisão. Para a OCDE (2017b), se uma certa quantidade de tensão ou preocupação é essencial para a motivação e o alto desempenho, muita pressão pode prejudicar o desempenho de um aluno e o seu bem-estar psicológico. Tanto os professores quanto os pais precisam encontrar maneiras de incentivar a motivação dos alunos em aprender e alcançar bons resultados sem gerar um medo excessivo de fracasso.

Na área de Ciências, verificou-se 14 incidências de atributos especificamente sobre essa área, sendo um deles apresentado como raiz. Os atributos foram: ST092Q01TA, JOYSCIE, ST129Q02TA, EPIST, ST131Q04NA e ST131Q06NA. O atributo ST092Q01TA se apresentou como raiz em 3 dos 5 países com melhor desempenho em Ciências no PISA 2015. Verificou-se que o fato de os alunos terem informação sobre os temas de cada atributo juntamente com o alto índice de confiança sobre abordagens científicas e a apreciação dele por essa área foi relevante para se alcançar os quartis que representam as médias mais altas em Ciências. Para a OCDE (2016b) a forma como a ciência é ensinada na escola pode influenciar não apenas se os alunos se saem bem nessa área, mas também se eles se interessam o suficiente sobre o assunto.

Verificou-se que os atributos relacionados a prática de atividade física tiveram 11 incidências nas 3 áreas de conhecimento. Os atributos sobre esse tema são: ST076Q11NA, ST078Q11NA e o atributo ST032Q01NA. Verificou-se que os alunos que praticam atividade física em algum momento do dia tenderam alcançar os quartis que representam as médias mais altas. Borgonov e Pál (2016)

afirmaram que a prática de esportes individuais e coletivos estão associados ao desenvolvimento físico, emocional, social, cognitivo em geral. Nesse sentido, Greeff *et al.* (2018) constataram que a prática de atividade física regular contínua durante várias semanas tem efeitos positivos no desempenho dos alunos. Para a OCDE (2018), envolver-se em atividade física moderada por pelo menos 60 minutos por dia está ligado, embora modestamente, a um maior desempenho acadêmico, e ressaltaram que os educadores e as escolas podem, portanto, reforçar o bem-estar dos alunos com atividade física moderada de qualidade e ajudando-os a permanecer fisicamente ativos, dentro e fora da escola.

Os atributos relacionados ao tempo de aprendizagem de cada disciplina tiveram 10 incidências nas 3 áreas de conhecimento. Na área de Matemática, essa questão é representada pelos atributos MMINS e ST059Q02TA; em Leitura, pelo atributo LMINS; e em Ciências, pelo atributo SMINS que tem conexão com o atributo ST063Q02NA (se o aluno teve aula de química). Verificou-se que os alunos que alcançaram as médias representadas nos quartis mais altos foram aqueles que tiveram mais aulas de Matemática, Leitura e Ciências durante a semana. Em Leitura, verificou-se que o fato de assistir aulas adicionais, além de aulas obrigatórias (HADDINST) também foi relevante para se obter um bom desempenho. Segundo a OCDE (2016b), combinando o número total de horas que os alunos passam aprendendo ou estudando dentro e fora da escola e suas pontuações em Matemática, Leitura e Ciências é possível ter uma ideia aproximada de como os alunos são eficientes em seu aprendizado. A análise do tempo de aprendizagem pode indicar a qualidade do sistema de ensino adotado e também pode ser indicativo de que, para ter sucesso acadêmico, os alunos precisam dedicar mais tempo ao aprendizado.

Sobre a questão do trabalho, verificou-se que o atributo ST076Q09NA (nos últimos dias, fez o serviço de casa ou cuidou de outros membros da família antes de ir para a escola) teve incidência somente na área de Matemática, e tem conexão com o atributo ST076Q10NA (Nos últimos dias, realizou trabalho remunerado antes de ir para a escola) e com o atributo ST078Q10NA (Nos últimos dias, realizou trabalho remunerado após sair da escola). No total, esses atributos tiveram 5 incidências nas 3 áreas de conhecimento. Constatou-se que a realização dessas atividades antes de ir para a escola afeta o desempenho dos alunos, tendo em vista que os alunos que tenderam alcançar os quartis que representam as médias mais altas nas 3 áreas de conhecimento são aqueles que não realizam serviços em casa, bem como não exercem atividade remunerada. Corroborando com a OCDE (2017b), que constatou que alunos que exercem trabalho remunerado ou que trabalham em casa tendem a ter uma pontuação mais baixa do que aqueles que não trabalham. Em média, a diferença de pontos no

desempenho científico entre os estudantes que trabalham no lar e os que não trabalham é de 13 pontos, enquanto a diferença é de 55 pontos entre os estudantes que trabalham por remuneração e aqueles que não trabalham. Essas descobertas sugerem que o desligamento da escola está correlacionado com o status de emprego dos alunos.

Os atributos CPSVALUE e ST082Q13NA, que são relacionados com a colaboração e trabalho em equipe, tiveram 3 incidências na área de Leitura. Verificou-se que a cooperação e trabalho em equipe atuaram de forma positiva no desempenho dos alunos que alcançaram os quartis que representam as médias mais altas dessa área. Fung, Hung e Lui (2018) constataram que a utilização de estratégias eficazes aplicados ao trabalho em grupo elevou o resultado de testes realizados pelos alunos. Além disso, foi possível identificar o aumento da construção conjunta de conhecimentos conceituais relacionados às ciências.

Verificou-se que na área de Matemática, apresentada na Seção 4.5.1, o atributo REPEAT (índice de reprovação) se apresentou como raiz, e teve mais 1 incidência somente nessa área de conhecimento. Constatou-se que os alunos que já repetiram de ano em algum nível tenderam a alcançar os quartis 1 e 2, que representaram as menores médias em Matemática. Bacchetto (2016) afirmou que os estudantes reprovados em algum ano, em média, apresentaram resultados em avaliações piores do que os estudantes não repetentes. O autor também discutiu a eficácia pedagógica deste mecanismo. Internacionalmente, a OCDE (2016c) constatou que a repetição de série é ineficaz na equalização do desempenho do aluno, uma vez que os alunos que são retidos tendem a sofrer perdas de aprendizagem em relação àqueles não retidos.

O atributo ISCEDD teve 2 incidências, sendo 1 na área de Matemática e outra na área de Ciências. Conforme já mencionado na Seção 4.5.1, o atributo ISCEDD indica o fluxo do programa de estudo do aluno, sendo as opções: educação acadêmica continuada, educação técnica avançada ou mercado de trabalho. Na intersecção dos resultados, verificou-se que esse atributo foi relevante para se alcançar as médias representadas nos quartis 3 e 4 nas duas áreas de conhecimento. Segundo a OCDE (2016b), os alunos dos programas de educação acadêmica continuada ou técnica avançada alcançaram 22 pontos a mais na avaliação científica do PISA 2015 do que os estudantes dos programas pré-profissionais ou para o mercado de trabalho. Nesse sentido, Burhan *et al.* (2017) afirmaram que as instituições educacionais deveriam ser capazes de melhorar compreensão do aluno sobre o mundo em que vivem, e capacitá-los a se tornarem participantes ativos na transformação da sociedade.

O atributo UNFAIRTEACHER (índice relacionado a imparcialidade do professor), que teve 1 incidência na área de Ciências, está relacionado com o atributo ST039Q01NA (frequência que o aluno teve a seguinte experiência: “Os professores me chamaram com menos frequência com que eles chamaram outros alunos”), que teve 1 incidência na área de Matemática. Nas duas áreas de conhecimento, verificou-se que esses atributos foram relevantes para se alcançar os quartis 3 e 4, que representaram as maiores médias. Para a OCDE (2017b), as relações negativas professor-aluno podem minar a confiança dos alunos e levar a maior ansiedade, afetando o desempenho acadêmico. Os autores afirmaram que os professores devem ser treinados em métodos básicos de observação, comunicação e intercultural, para que possam levar mais em conta as necessidades individuais dos alunos, e devem ser incentivados a colaborar e trocar informações sobre as dificuldades e pontos fortes dos alunos com seus colegas, para que possam coletivamente encontrar a melhor abordagem para fazer com que os alunos se sintam parte da comunidade escolar.

O atributo ST076Q07NA (nos últimos dias, encontrou-se com amigos ou conversou com amigos pelo telefone antes de ir para a escola) teve incidência somente na área de Matemática. Verificou-se que os alunos que responderam não a esse questionamento alcançaram os quartis que representam as médias mais altas. Para a OCDE (2017c), encontrar amigos, conversar com amigos no telefone e usar a *Internet*, o bate-papo ou as redes sociais são maneiras de desenvolver e nutrir relacionamentos com os outros. Portanto, pode parecer surpreendente que essas atividades estejam associadas a uma diferença no desempenho dos alunos e também como eles valorizam o trabalho em equipe. No entanto, uma explicação para essas relações não pode ser determinada a partir dos dados do PISA. Sobre esse tema, Borgonov e Pál (2016) afirmaram que as interações com os amigos podem facilitar o desenvolvimento social dos alunos, a motivação acadêmica e o desempenho.

O atributo ST078Q02NA (nos últimos dias, estudou ou fez tarefa de casa após sair da escola) teve incidência somente na área de Matemática. Para Oliveira e Casagrande (2017), a atividade de casa é um instrumento pedagógico fundamental para o processo de ensino-aprendizagem dos alunos e deve ser compreendida como uma construção da autonomia na aprendizagem. Para a OCDE (2016b) fazer tarefa de casa pode ajudar o aluno a identificar e aplicar o conteúdo que aprendeu e fornece estímulo adicional para um alto desempenho estudantil.

O atributo ST001D01T (série/ano que o aluno se encontra) teve 1 incidência somente na área de Leitura. Verificou-se que esse atributo foi relevante para se alcançar os quartis 2 e 3. Segundo a OCDE (2016b) a série em que os alunos estavam matriculados no momento em que realizaram a avaliação do PISA depende em grande parte de três fatores: idade, a idade em que iniciaram o ensino

primário e, sobretudo, se repetiram ou não uma nota. Na média, 28% da variação no nível de série dos alunos é explicada pelo fato de terem repetido ou não uma série na educação primária ou secundária, 13% na idade dos alunos (alguns alunos estão matriculados em graus mais altos / mais baixos apenas porque nasceram mais cedo ou mais tarde) e 4% pela idade em que ingressaram no ensino primário.

O atributo PA006Q02TA (se a instituição ter uma boa reputação é uma razão para escolher uma escola para o filho) teve 1 incidência somente na área de Matemática. Verificou-se que os alunos cujos os pais afirmaram que essa questão é muito importante tenderam a alcançar médias maiores do que aqueles alunos que os pais não consideraram essa questão importante. MacLeod e Urquiola (2009) afirmaram que a admissão em escolas que detêm uma boa reputação fazem com que haja uma sinalização de que o aluno possui alta capacidade intelectual, o que sugere uma maior probabilidade de se ter uma boa renda futura. Atamturk (2017) identificou que os pais consideram a reputação da escola como uma ferramenta em potencial para uma construção intelectual de qualidade para seus filhos.

4.7. Conclusão

Ao se aplicar mineração de dados na base do PISA foi possível identificar e analisar, por meio das árvores de decisão geradas, os padrões que expliquem o sucesso dos países que se destacaram na avaliação, atendendo assim o objetivo da pesquisa. Esses padrões estão relacionados aos fatores socioeconômicos, aspiração do nível educacional do aluno, uso de TIC's, motivação e ansiedade, abordagens científicas, prática de atividade de física, tempo de aprendizagem em cada disciplina, trabalho remunerado, reprovação, cooperação e trabalho em equipe, programa de estudo do aluno, imparcialidade do professor, interações com amigos antes de ir para a escola, tarefa de casa, série/ano que o aluno se encontra e boa reputação da escola.

Não obstante as taxas de acertos obtidas não terem ultrapassado 65%, os resultados são considerados satisfatórios por permitirem relacionar os atributos e realizar análises que de fato descreveram os padrões encontrados. As análises apontaram que os atributos relacionados aos fatores socioeconômicos e a aspiração educacional do aluno tiveram uma maior importância no desempenho dos alunos nas 3 áreas de conhecimento, afetando diretamente nos resultados da avaliação. Nesse sentido, Paulino (2013) apontou a existência de uma relação entre esses dois atributos, ressaltando que os alunos com nível socioeconômico elevado tendem a ter aspirações educacionais elevadas. Utilizando dados internacionais, tais resultados corroboraram também com vários estudos presentes

na literatura que indicaram que o nível socioeconômico é considerado muito importante na área educacional em razão da sua forte correlação com o desempenho dos estudantes.

A análise realizada neste estudo poderá direcionar os professores e gestores educacionais para novos procedimentos nos processos de ensino e aprendizagem, contribuindo efetivamente para a formação de estudantes, fortalecendo os atos de aprender, de ensinar e de formar profissionais e cidadãos. No contexto das políticas públicas, espera-se que essa pesquisa contribua para o desenho de políticas, programas e ações que melhorem as condições de funcionamento da escola ao tempo que forcem mais recursos para a política educacional.

Para trabalhos futuros, recomenda-se a aplicação de DM na base de dados do PISA para analisar os países que tiveram baixo desempenho nas 3 áreas de conhecimento e realizar a comparação com os resultados desta pesquisa, no sentido de oferecer informações que contribuam para a elaboração de políticas estudantis.

Referências Bibliográficas

AKSU, G.; GÜZELLER, C. O. Classification of PISA 2012 Mathematical Literacy Scores Using Decision-Tree Method: Turkey Sampling. **Ted Egitim ve Bilim**, v. 41, n. 185, 9 jun. 2016.

ALVES, F. DE A.; SILVA FILHO, O. C. **O Efeito da Escola e os Determinantes do Rendimento Escolar: Uma análise dos resultados dos estudantes brasileiros nas últimas três edições do PISA**. In: 45º ENCONTRO NACIONAL DE ECONOMIA. Natal/RN: ANPEC, 2017.

ATAMTURK, N. Parental construction of school quality when making school choice. **Quality & Quantity**, 2017.

AVSAR, A. S.; TAVSANCIL, E. Determination variables which define Turkish students' PISA 2009 reading literacy with chaid analysis. **Pakistan Journal of Statistics**, v. 30, n. 6, p. 1179–1185, 2014.

BABY, A. M. Pedagogue Performance Assessment (PPA) using Data mining Techniques. **IOP Conference Series: Materials Science and Engineering**, v. 396, p. 012024, 29 ago. 2018.

BACCHETTO, J. G. O Pisa e o custo da repetência no Fundeb. **Ensaio: Avaliação e Políticas Públicas em Educação**, v. 24, n. 91, p. 424–444, jun. 2016.

BADRI, M. *et al.* An analytic hierarchy process for school quality and inspection: Model development and application. **International Journal of Educational Management**, v. 30, n. 3, p. 437–459, 2016.

BORGONOV, F.; PÁL, J. **A Framework for the Analysis of Student Well-Being in the PISA 2015 Study: Being 15 In 2015**. Paris: OECD Publishing, 2016.

- BURHAN, N. A. S. *et al.* Why are cognitive abilities of children so different across countries? The link between major socioeconomic factors and PISA test scores. **Personality and Individual Differences**, v. 105, p. 95–106, jan. 2017.
- CHANG, D.-F.; CHEN, C.-C. CLUSTER ANALYSIS FOR STUDENT PERFORMANCE IN PISA2015 AMONG OECD ECONOMIES. **ICIC Express Letters**, v. 9, n. 11, p. 8, 2018.
- CICEO, G. Cultural capital and reading performance of the Romanian students in post-communist Romania. Can the gaps still be recovered? **On-line Journal Modelling the New Europe**, v. 24, p. 3–22, 15 jan. 2018.
- CORDEIRO, R. G.; HORA, H. R. M.; MUSSA, M. S. **Comportamento de estudantes evadidos de cursos técnicos: Um estudo utilizando técnicas de mineração de dados**. Memórias de la Octava Conferencia Iberoamericana de Complejidad, Informática y Cibernética. Anais... In: Octava Conferencia Iberoamericana de Complejidad, Informática y Cibernética: CICIC 2018. Orlando, Florida: 2018.
- DORN, E. *et al.* **Fatores que influenciam o sucesso escolar na América Latina**. McKinsey & Company, 2017.
- FAYYAD, U.; PIATETSKY-SHAPIRO, G.; SMYTH, P. From data mining to knowledge discovery in databases. **AI magazine**, v. 17, n. 3, p. 37, 1996.
- FERNANDES, R. A universalização da avaliação e a criação do Ideb: pressupostos e perspectivas. **Em Aberto**, v. 29, n. 96, 2016.
- FERREIRA, G. **Investigação acerca dos fatores determinantes para a conclusão do Ensino Fundamental utilizando Mineração de Dados Educacionais no Censo Escolar da Educação Básica do INEP 2014**. Anais dos Workshops do CBIE 2015. **Anais...** In: CONGRESSO BRASILEIRO DE INFORMÁTICA NA EDUCAÇÃO 2015. Maceió, AL: 2015. Disponível em: <<http://br-ie.org/pub/index.php/wcbie/article/view/6211>>. Acesso em: 14 fev. 2018
- FUNG, D.; HUNG, V.; LUI, W. Enhancing Science Learning Through the Introduction of Effective Group Work in Hong Kong Secondary Classrooms. **International Journal of Science and Mathematics Education**, v. 16, n. 7, p. 1291–1314, out. 2018.
- GAMAZO, A. *et al.* Evaluación de factores relacionados con la eficacia escolar en PISA 2015. Un análisis multinivel. **Revista de Educación**, v. 379, p. 56–78, 2018.
- GIL, A. C. **Como elaborar projetos de pesquisa**. São Paulo: Atlas, 2009.
- GONÇALVES, S. S. B. **Envolvimento do Aluno na Escola, Percepção de Apoio Familiar e Desempenho Escolar**. Dissertação (Mestrado) - Portugal: Universidade da Madeira, 2017.
- GREEFF, J. W. *et al.* Effects of physical activity on executive functions, attention and academic performance in preadolescent children: a meta-analysis. **Journal of Science and Medicine in Sport**, v. 21, n. 5, p. 501–507, maio 2018.

HUSSAIN, S. *et al.* Educational Data Mining and Analysis of Students' Academic Performance Using WEKA. **Indonesian Journal of Electrical Engineering and Computer Science**, v. 9, n. 2, 2018.

INEP. **Brasil no PISA 2015 : análises e reflexões sobre o desempenho dos estudantes brasileiros / OCDE-Organização para a Cooperação e Desenvolvimento Econômico**. Brasília/DF: Fundação Santillana, 2016.

KASTHURIARACHCHI, K. T. S.; LIYANAGE, S. R.; BHATT, C. M. A Data Mining Approach to Identify the Factors Affecting the Academic Success of Tertiary Students in Sri Lanka. In: CABALLÉ, S.; CONESA, J. (Eds.). . **Software Data Engineering for Network eLearning Environments**. Cham: Springer International Publishing, 2018. v. 11p. 179–197.

LIU, X.; WHITFORD, M. Opportunities-to-Learn at Home: Profiles of Students With and Without Reaching Science Proficiency. **Journal of Science Education and Technology**, v. 20, n. 4, p. 375–387, ago. 2011.

MACLEOD, W. B.; URQUIOLA, M. **Anti-lemons: school reputation and educational quality**. National Bureau of Economic Research, 2009.

MARQUES, L. R. Qualidade educacional: sentidos hegemônicos em escolas com bom desempenho em avaliações de larga escala. **Jornal de Políticas Educacionais**, v. 11. n.º 23, 2017.

MOURA, I. A.; SADOYAMA, A. DOS S. P. Qualidade dos serviços educacionais prestados na Educação Infantil: novos olhares para práticas gestoras. **Revista EDaPECI**, v. 17, n. 1, p. 218–233, 2017.

MUÑOZ, J.; CHARRO, E. La interpretación de datos y pruebas científicas vistas desde los ítems liberados de PISA. **Revista Eureka sobre Enseñanza y Divulgación de las Ciencias**, Universidad de Cádiz. v. 15, n. 2, 2018.

NICOLETTI, M. C.; MARQUES, M.; GUIMARAES, M. P. **A data mining approach for forecasting students' performance**. 13th Iberian Conference on Information Systems and Technologies (CISTI). **Anais...** In: 13Th Iberian Conference on Information Systems and Technologies (Cisti). Cáceres: IEEE, jun. 2018. Disponível em: <<https://ieeexplore.ieee.org/document/8399389/>>. Acesso em: 10 dez. 2018

OCDE. **Resumo de resultados nacionais do PISA 2015 - Brasil**. Paris: OCDE *Publishing*, 2016a. Disponível em: <<http://www.oecd.org/pisa/pisa-2015-Brazil.pdf>>. Acesso em: 10 fev. 2018.

OCDE. **PISA 2015 Results (Volume II)**. Paris: OCDE *Publishing*, 2016b. Disponível em: <<http://www.oecd.org/education/pisa-2015-results-volume-ii-9789264267510-en.htm>>.

OCDE. **PISA 2015 Results (Volume I)**. Paris: OCDE *Publishing*, 6 dez. 2016c. Disponível em: <http://www.oecd-ilibrary.org/education/pisa-2015-results-volume-i_9789264266490-en>. Acesso em: 10 fev. 2018.

OCDE. **PISA 2015 technical report**. Paris: OCDE *Publishing*, 2017a. Disponível em: <<http://www.oecd.org/pisa/data/2015-technical-report/>>. Acesso em: 11 fev. 2018.

OCDE. **PISA 2015 Results (Volume III): Students' Well-Being**. Paris: OCDE *Publishing*, 2017b. Disponível em: <<https://www.oecd.org/pisa/PISA-2015-Results-Students-Well-being-Volume-III-Overview.pdf>>.

OCDE. **PISA 2015 Results (Volume V)**. Paris: OCDE *Publishing*, 2017c. Disponível em: <https://www.oecd-ilibrary.org/education/pisa-2015-results-volume-v/student-activities-school-practices-and-collaboration_9789264285521-11-en>. Acesso em: 9 dez. 2018.

OCDE. **How is participation in sports related to students' performance and well-being?** Paris: OCDE *Publishing*, 2018. Disponível em: <https://www.oecd-ilibrary.org/education/how-is-participation-in-sports-related-to-students-performance-and-well-being_e124db26-en>. Acesso em: 9 dez. 2018.

OLIVEIRA, G. D.; CASAGRANDE, S. Tarefa de casa: vilã ou protagonista do processo de ensino e aprendizagem. **Revista Saberes Pedagógicos**, v. 1, n. 1, p. 82, 31 jan. 2017.

OLIVEIRA, Q. C. DA S.; COELHO, D.; CASTANHA, A. P. Considerações sobre as avaliações em larga escala no Brasil e o papel dos organismos internacionais: eficiência e produtividade x qualidade. **Revista eletrônica de Política e Gestão Educacional**, v. 19, 2015.

PAULINO, D. S. R. **Autopercepções, contexto familiar e aspirações vocacionais de alunos do 2º ciclo**. Dissertação (Mestrado) - Lisboa: Universidade de Lisboa (Faculdade de Psicologia), 2013.

QIAO, X.; JIAO, H. Data Mining Techniques in Analyzing Process Data: A Didactic. **Frontiers in Psychology**, v. 9, 23 nov. 2018.

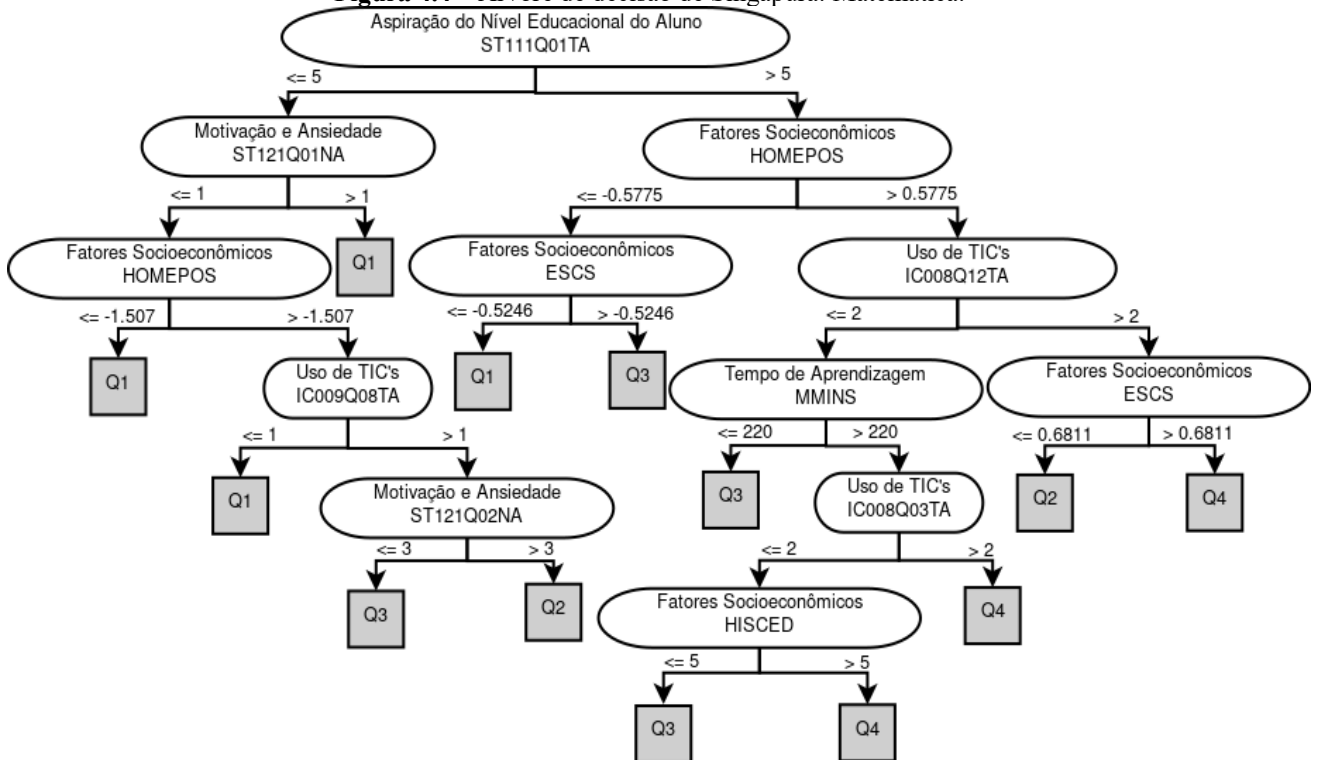
QUINLAN, J. R. Simplifying Decision Trees. **International Journal of Man-Machine Studies**, v. 27, n. 3, p. 221–234, 1987.

SALAH, N. B.; SAADI, I. B. **Fuzzy AHP for learning service selection in contextaware ubiquitous learning systems**. . In: 13TH IEEE INTERNATIONAL CONFERENCE ON UBIQUITOUS INTELLIGENCE AND COMPUTING. Tunis, Tunísia: Department of Computer Science, National Higher Engineering School of Tunis, 2017

SIMON, A.; CAZELLA, S. **Mineração de Dados Educacionais nos Resultados do ENEM de 2015**. Anais dos Workshops do CBIE 2017. **Anais...** In: Congresso Brasileiro de Informática na Educação. Recife, PE: 2017. Disponível em: <<http://www.br-ie.org/pub/index.php/wcbie/article/view/7461>>. Acesso em: 14 fev. 2018

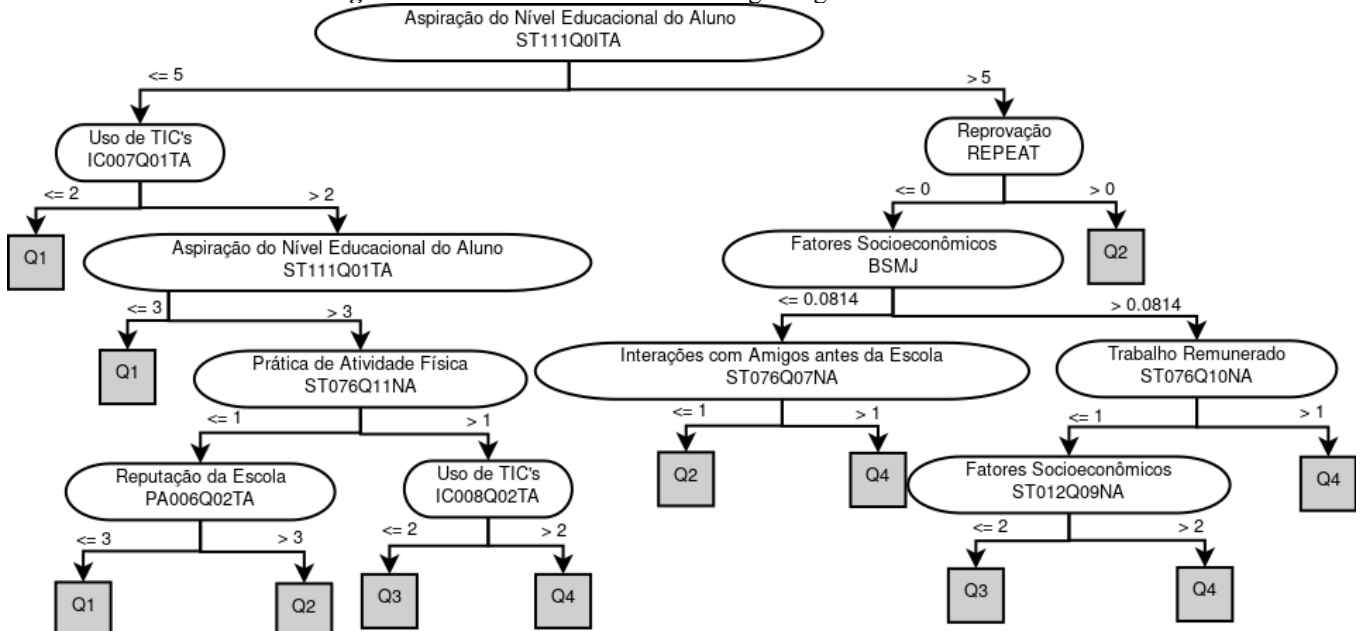
Apêndice A

Figura 4.4 – Árvore de decisão de Singapura: Matemática.



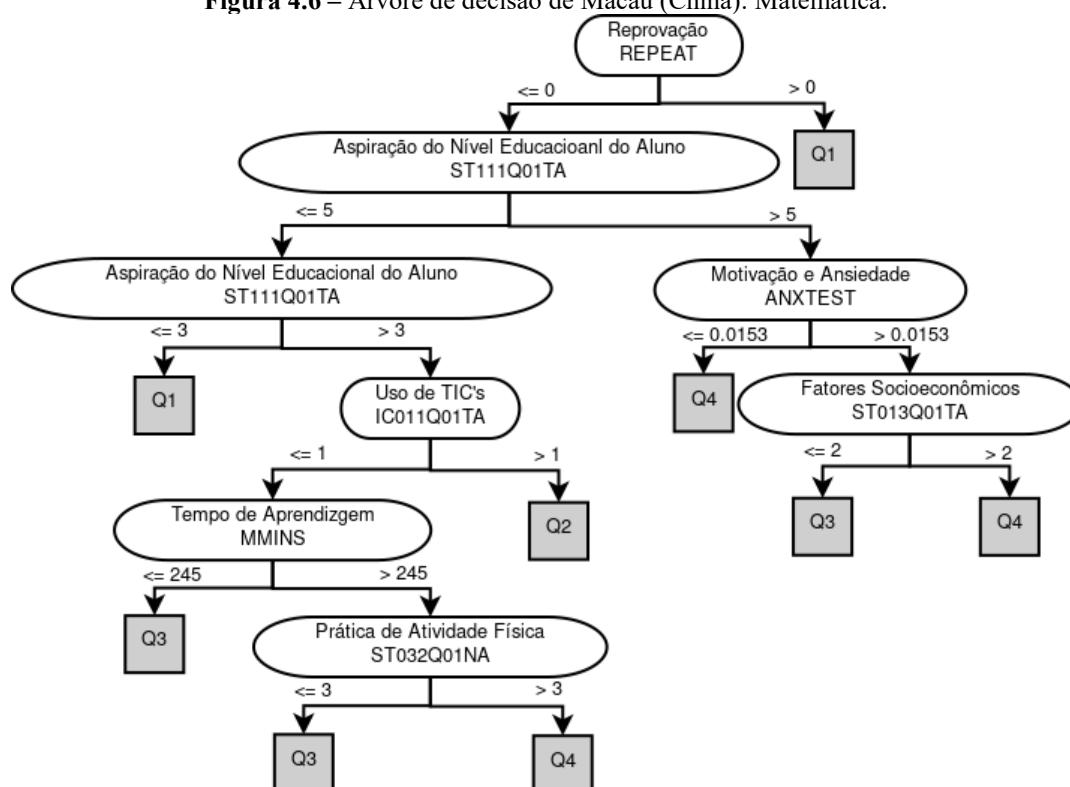
Fonte: Elaboração própria (2018).

Figura 4.5 – Árvore de decisão de Hong Kong: Matemática.



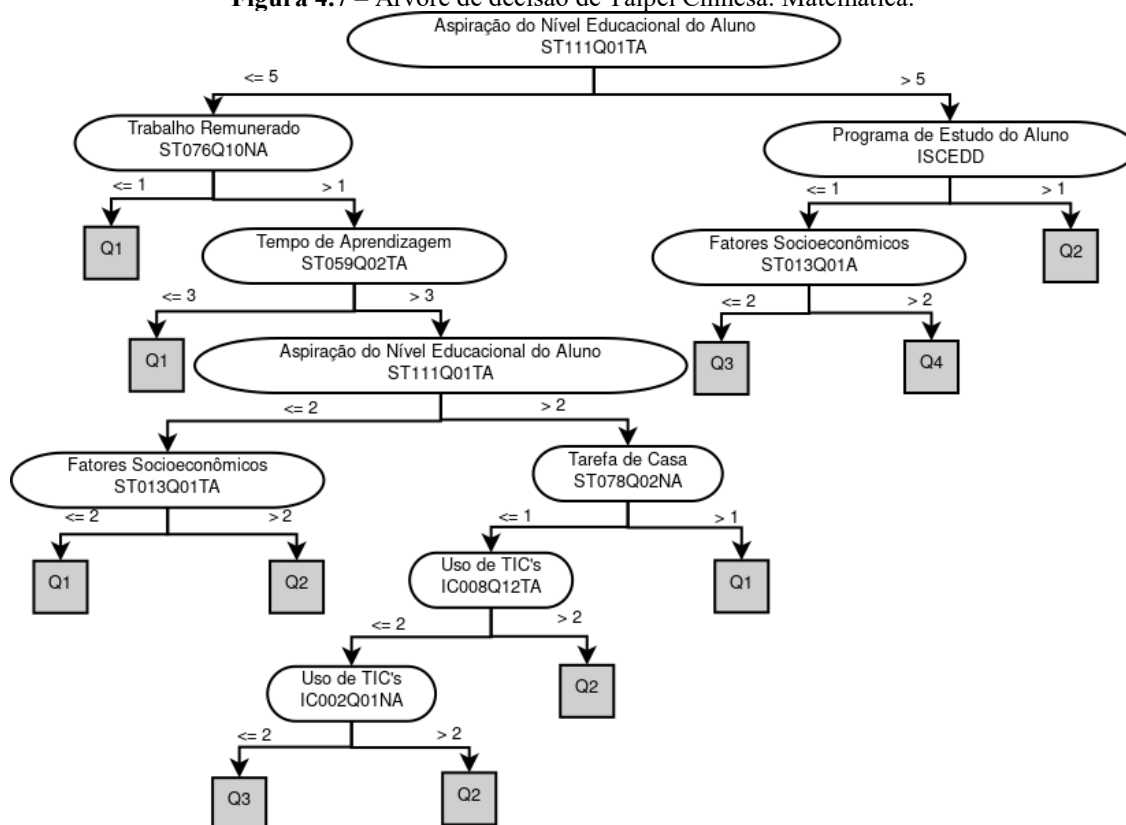
Fonte: Elaboração própria (2018).

Figura 4.6 – Árvore de decisão de Macau (China): Matemática.



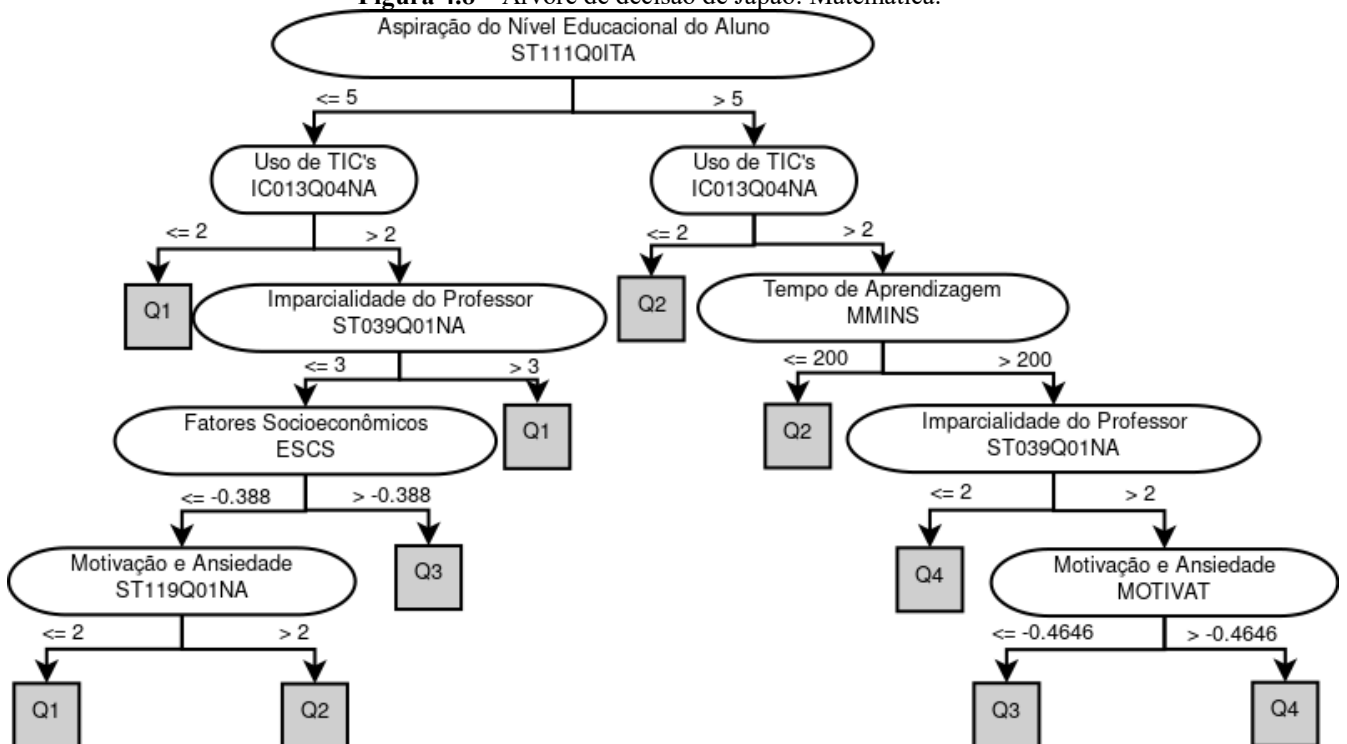
Fonte: Elaboração própria (2018).

Figura 4.7 – Árvore de decisão de Taipei Chinesa: Matemática.



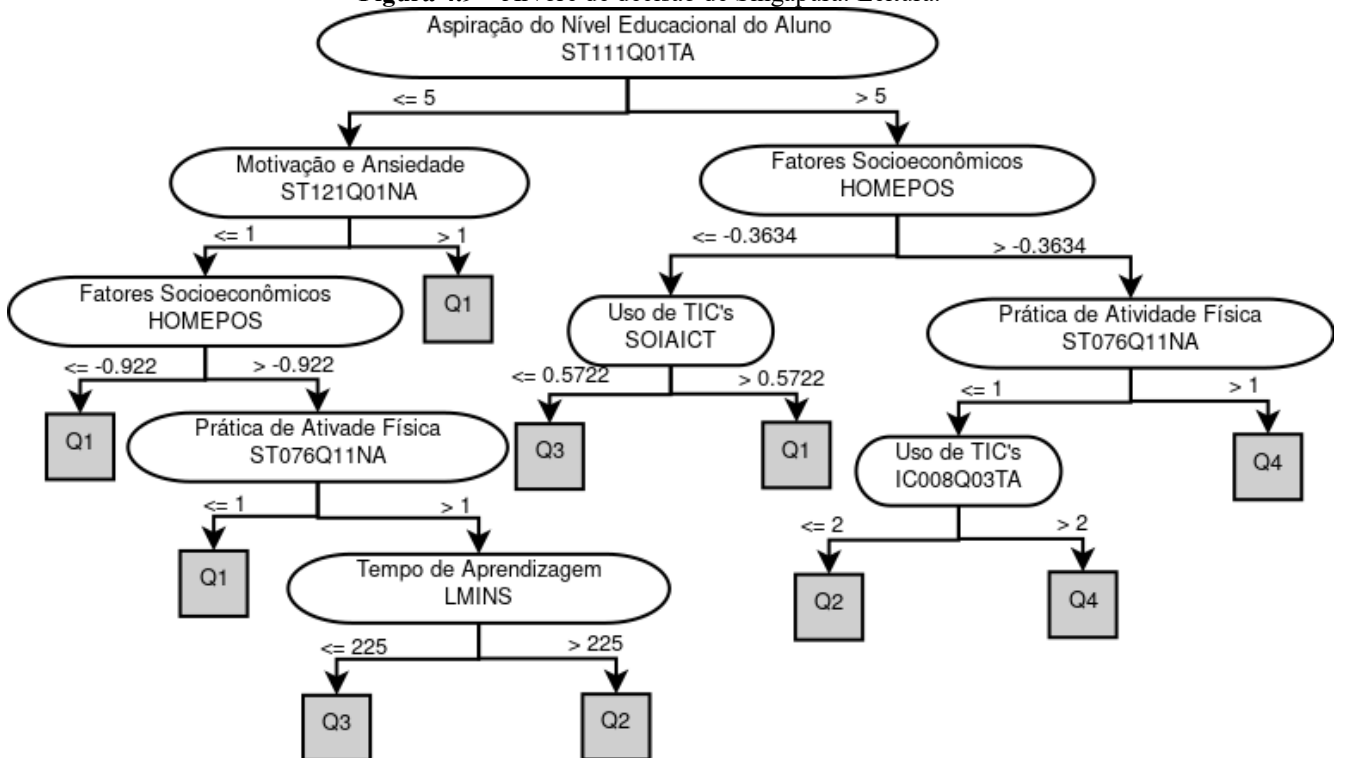
Fonte: Elaboração própria (2018).

Figura 4.8 – Árvore de decisão de Japão: Matemática.



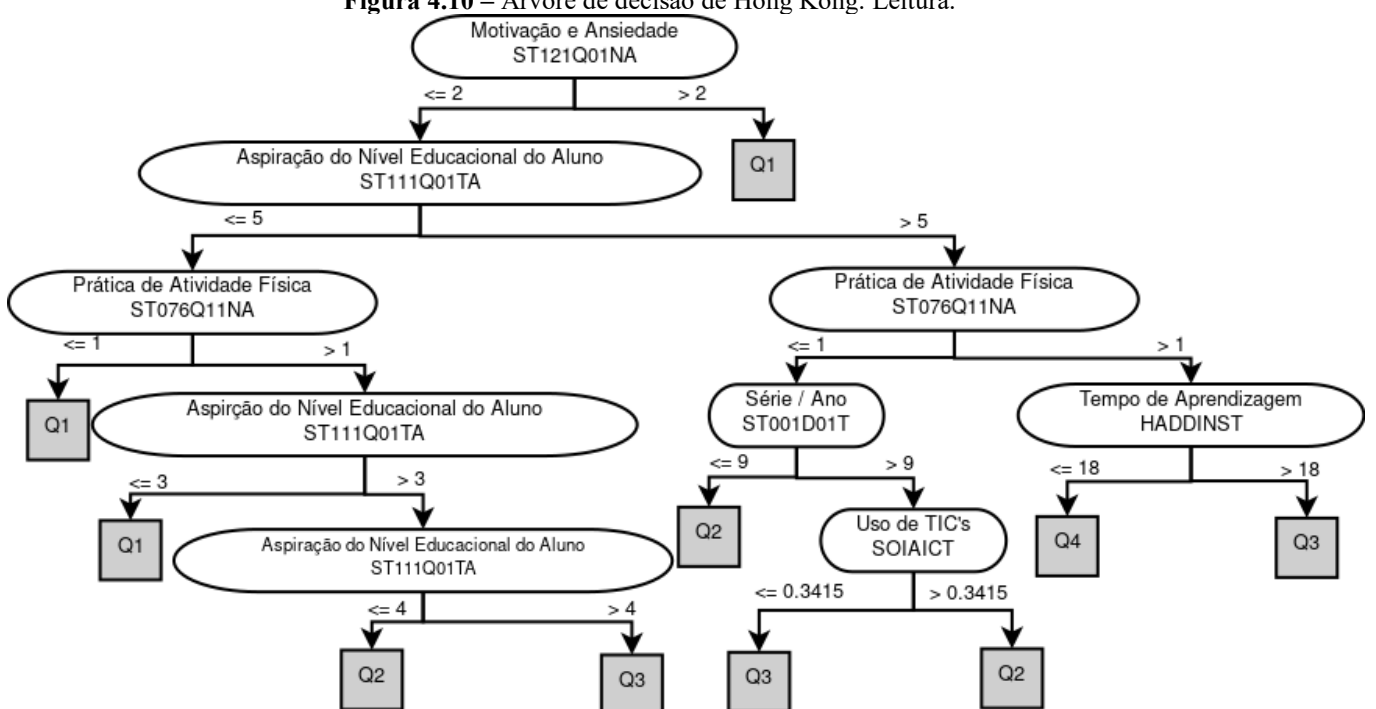
Fonte: Elaboração própria (2018).

Figura 4.9 – Árvore de decisão de Singapura: Leitura.



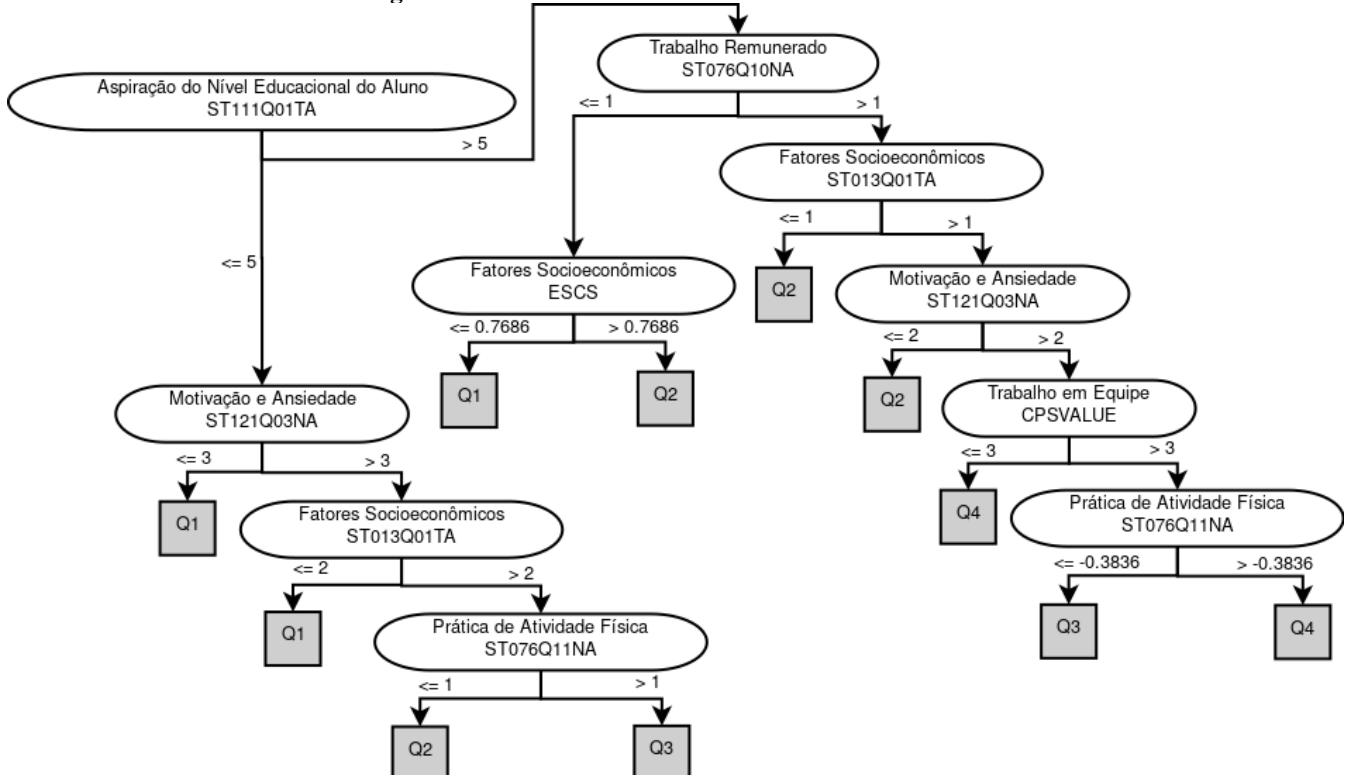
Fonte: Elaboração própria (2018).

Figura 4.10 – Árvore de decisão de Hong Kong: Leitura.



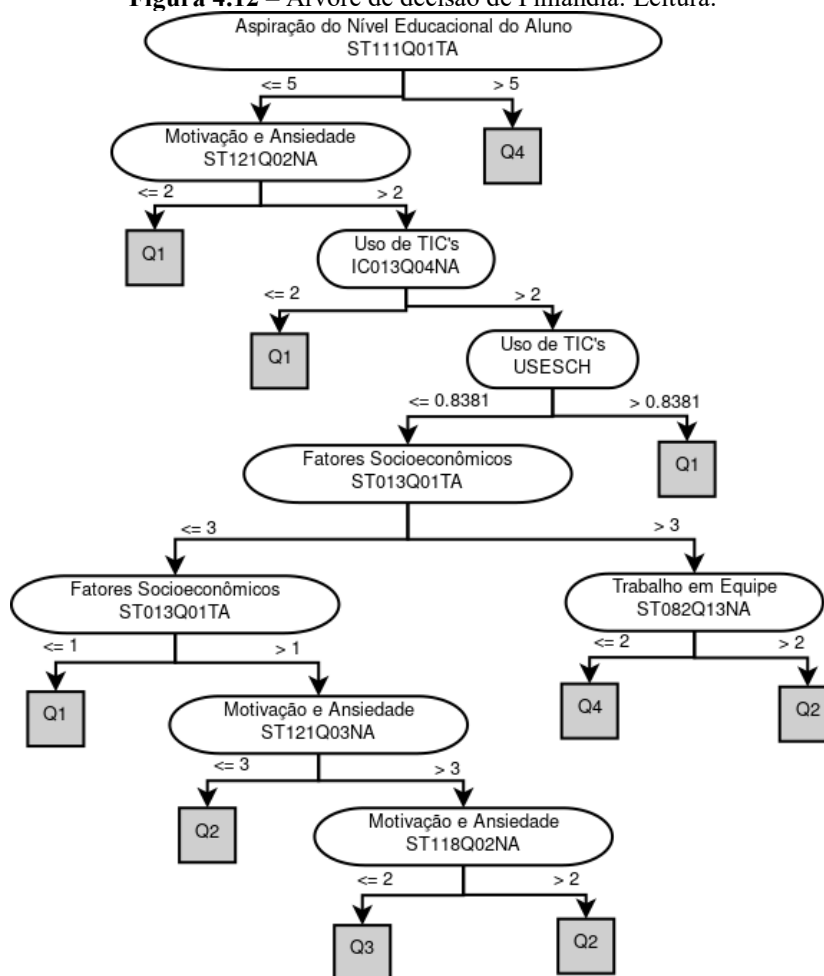
Fonte: Elaboração própria (2018).

Figura 4.11 – Árvore de decisão de Canadá: Leitura.



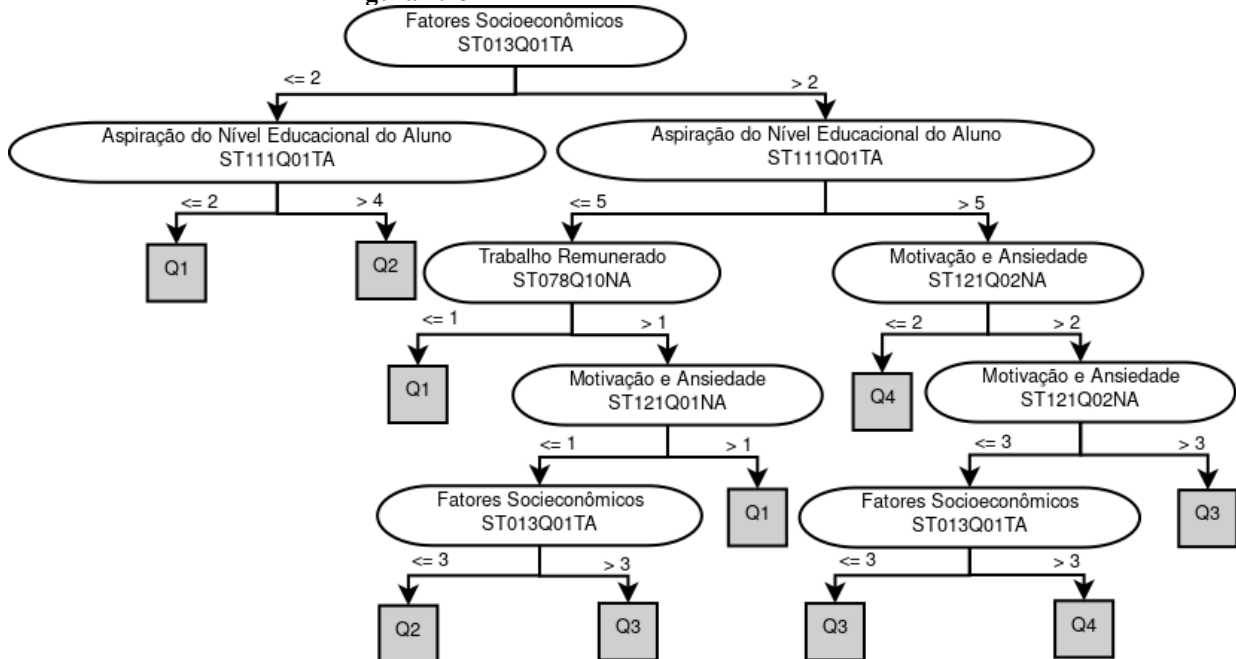
Fonte: Elaboração própria (2018).

Figura 4.12 – Árvore de decisão de Finlândia: Leitura.



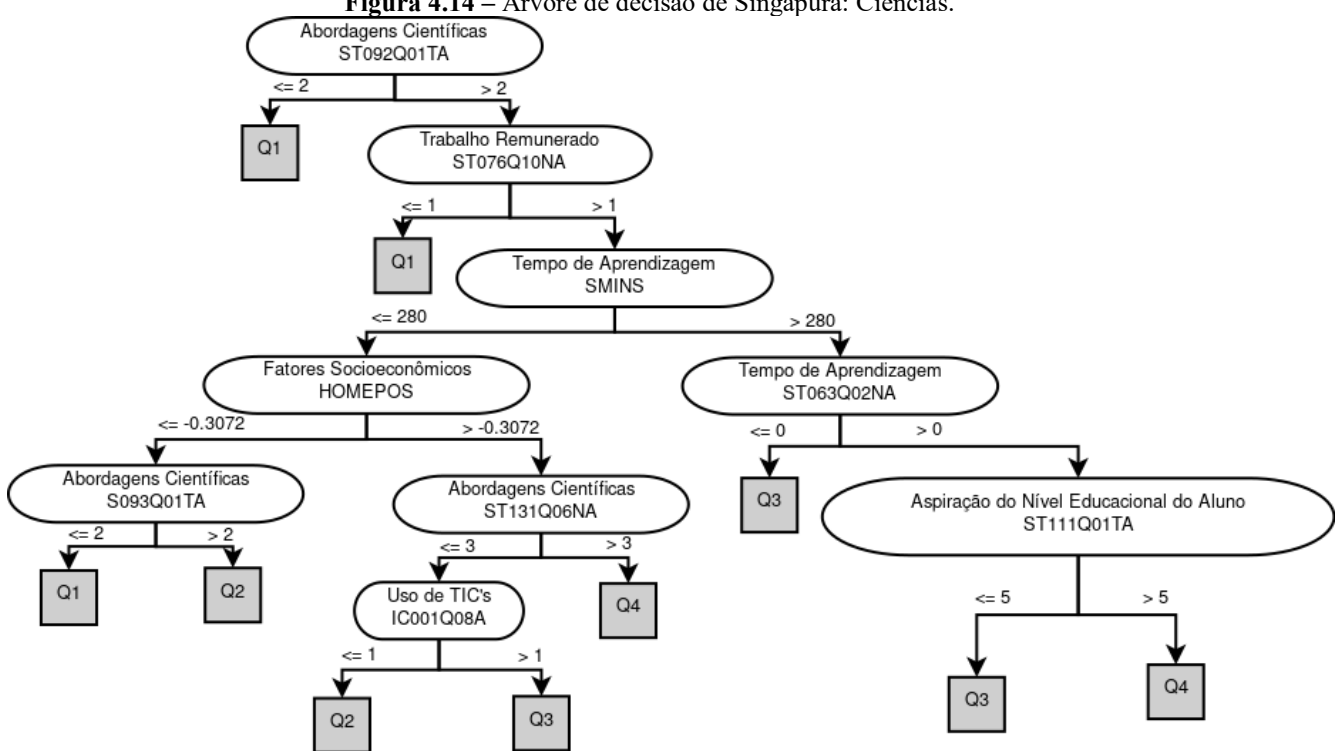
Fonte: Elaboração própria (2018).

Figura 4.13 – Árvore de decisão de Irlanda: Leitura.



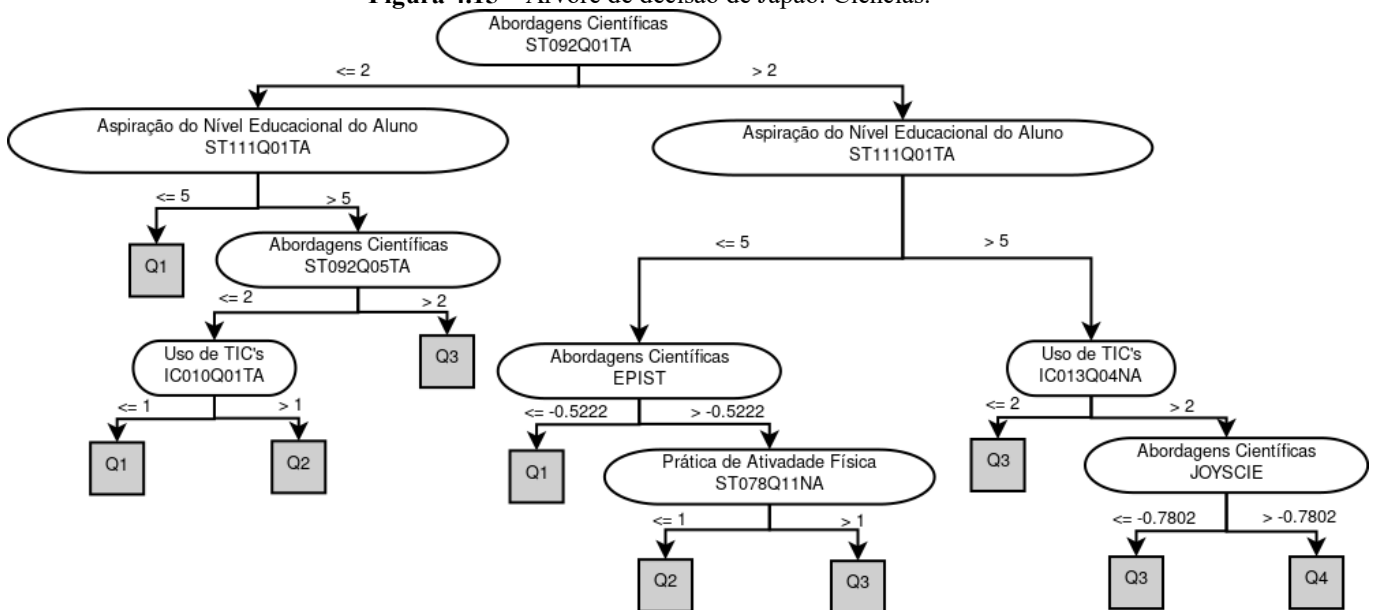
Fonte: Elaboração própria (2018).

Figura 4.14 – Árvore de decisão de Singapura: Ciências.



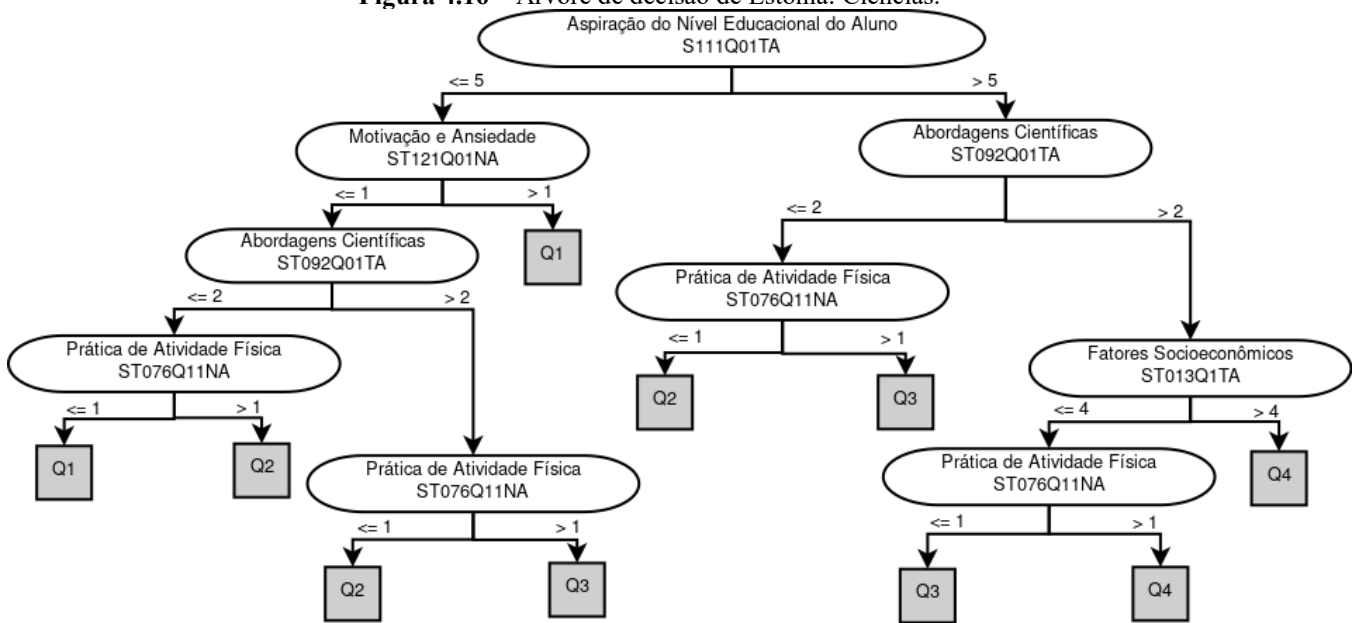
Fonte: Elaboração própria (2018).

Figura 4.15 – Árvore de decisão de Japão: Ciências.



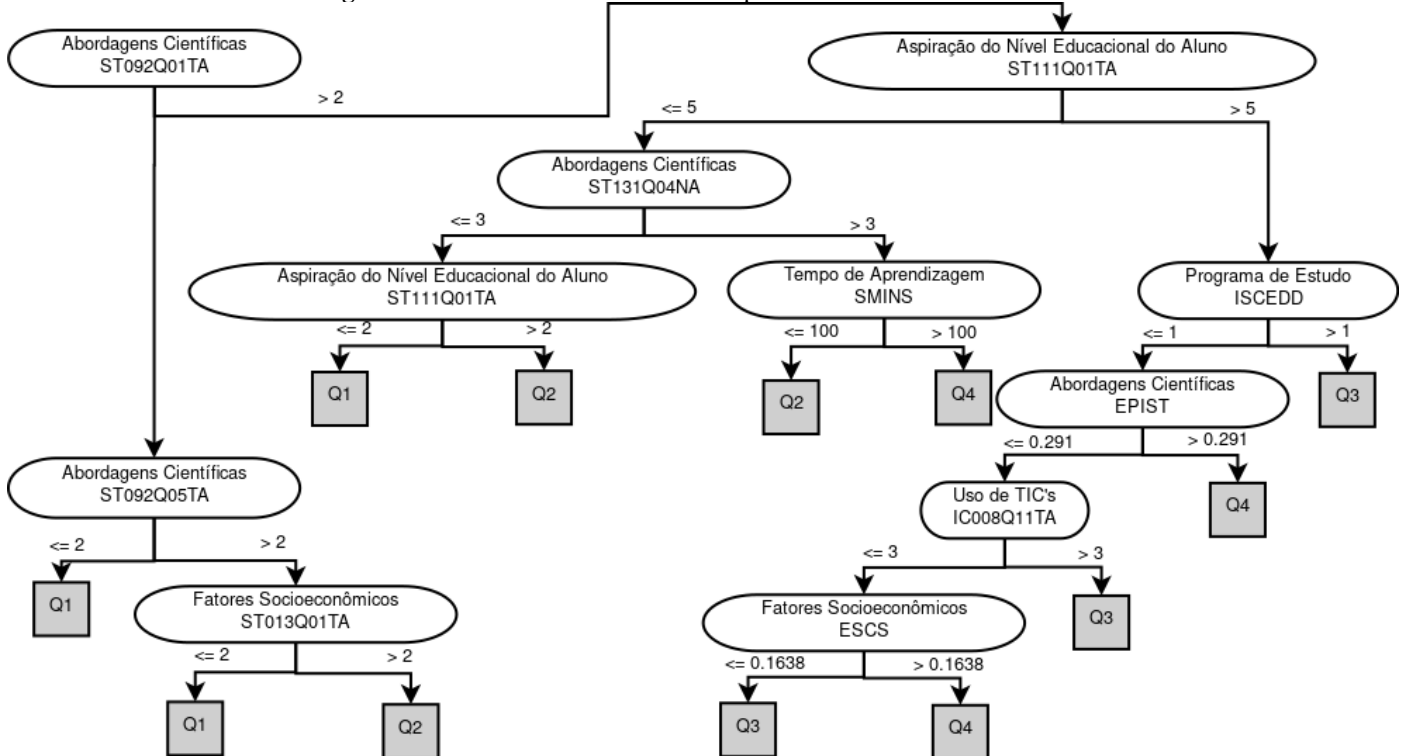
Fonte: Elaboração própria (2018).

Figura 4.16 – Árvore de decisão de Estônia: Ciências.



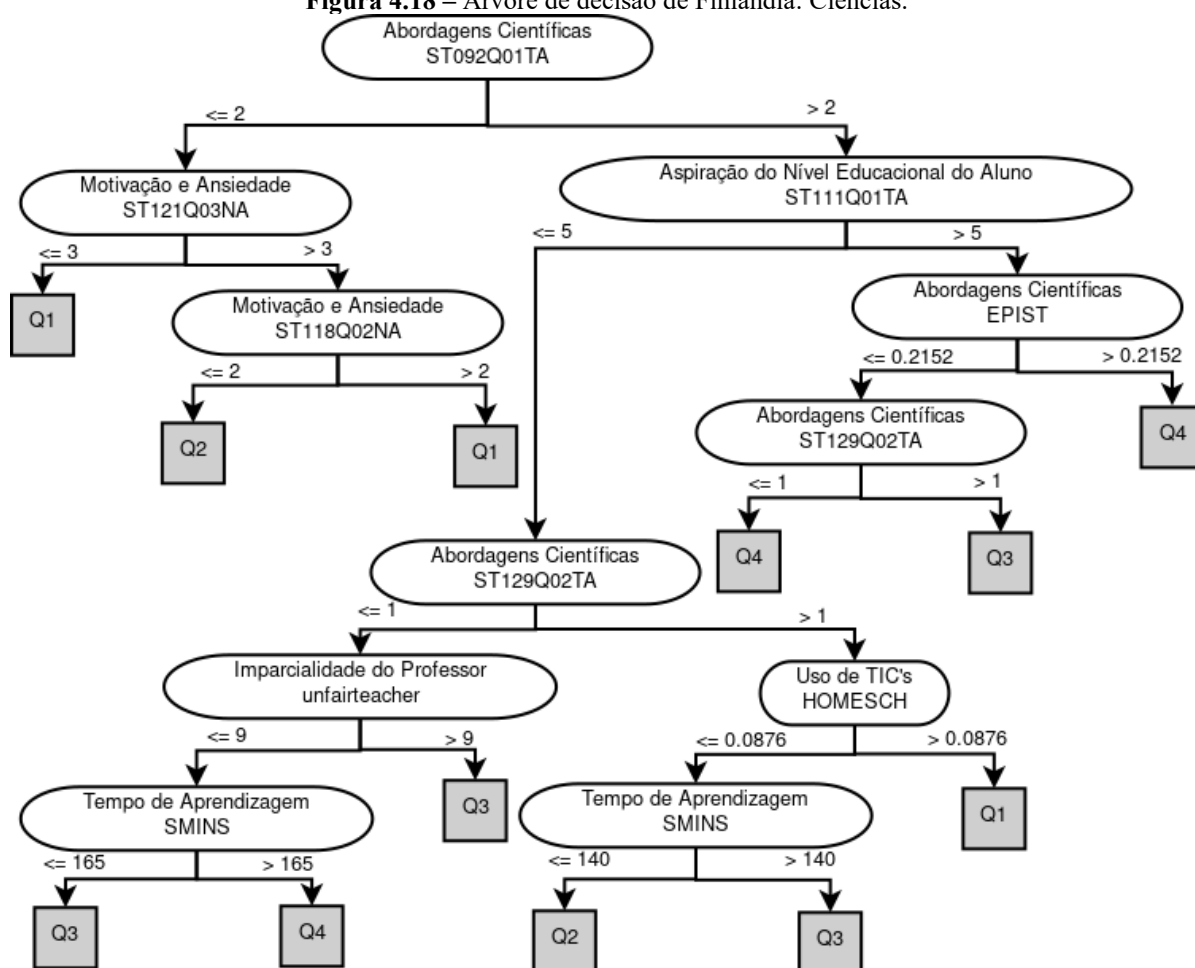
Fonte: Elaboração própria (2018).

Figura 4.17 – Árvore de decisão de Taipei Chinesa: Ciências.



Fonte: Elaboração própria (2018).

Figura 4.18 – Árvore de decisão de Finlândia: Ciências.



Fonte: Elaboração própria (2018).

5. CONSIDERAÇÕES FINAIS

A mineração de dados tem sido empregada para diferentes formas dentro da área da educação, sendo uma das melhores ferramentas inteligentes usadas para verificar o desempenho dos alunos (CHANG; CHEN, 2018). Este estudo teve como objetivo extrair conhecimento da base dados do PISA, utilizando técnicas de DM, de forma a se analisar quais são os fatores que foram relevantes para o sucesso educacional dos países que se destacaram na avaliação.

Quanto aos resultados, inicialmente, a pesquisa bibliométrica contribuiu para o conhecimento de um panorama mais abrangente sobre a literatura científica relacionada ao tema. Foi identificado que os estudos envolvendo mineração de dados na área educacional utilizando a base de dados do PISA são recentes e têm evoluído ao longo dos últimos anos, sendo notável que é uma área com grande potencial de ser explorada.

Na aplicação do método, foi possível identificar os atributos que foram relevantes nos países que se destacaram na avaliação, sendo possível mostrar a relação entre eles. Ressalta-se que a metodologia adotada neste estudo pode ser facilmente generalizada para outros algoritmos de mineração de dados.

Acredita-se que esta pesquisa irá colaborar direta e positivamente nas decisões dos professores e gestores, com o intuito de trazer melhorias para o ensino, tendo em vista que a avaliação do PISA constitui um recurso valioso para pensar, agir e legitimar as decisões.

REFERÊNCIAS BIBLIOGRÁFICAS

BARONE, R. E. M. Educação e políticas públicas: questões para o debate. **Boletim Técnico do Senac**, v. 30, n. 3, p. 57–71, 2017.

CHANG, D.-F.; CHEN, C.-C. CLUSTER ANALYSIS FOR STUDENT PERFORMANCE IN PISA2015 AMONG OECD ECONOMIES. **ICIC Express Letters**, v. 9, n. 11, p. 8, 2018.

DORN, E. *et al.* **Fatores que influenciam o sucesso escolar na América Latina**. McKinsey & Company, 2017.

FERRER, A. T. PISA in Spain: Expectations, impact and debate. **European Journal of Education**, v. 52, n. 2, p. 184–191, 2017.

INEP. **Brasil no PISA 2015 : análises e reflexões sobre o desempenho dos estudantes brasileiros / OCDE-Organização para a Cooperação e Desenvolvimento Econômico**. Brasília/DF: Fundação Santillana, 2016.

MARTÍNEZ, J. I. M.; CHARRO, E. La interpretación de datos y pruebas científicas vistas desde los ítems liberados de PISA. **Revista Eureka sobre enseñanza y divulgación de las ciencias.**, v. 15, n. 2, p. 1–20, 2018.

MURI, A. F.; ORTIGÃO, M. I. R. The Brazilian scientific literacy in the PISA. **Journal of Science Education**, v. 18, p. 80–84, 2017.

OCDE. **FT Sampling Guidelines**. Paris: OECD Publishing, 2015. Disponível em: <<https://www.oecd.org/pisa/pisaproducts/PISA2015FT-SamplingGuidelines.pdf>>. Acesso em: 10 dez. 2018.

OCDE. **Resumo de resultados nacionais do PISA 2015 - Brasil**. Paris: OCDE Publishing, 2016a. Disponível em: <<http://www.oecd.org/pisa/pisa-2015-Brazil.pdf>>. Acesso em: 10 dez. 2018.

OCDE. **PISA 2015 Results (Volume I)**. Paris: OCDE Publishing, 6 dez. 2016b. Disponível em: <http://www.oecd-ilibrary.org/education/pisa-2015-results-volume-i_9789264266490-en>. Acesso em: 10 fev. 2018.

PEJIC, A.; MOLCER, P. S. **Relationship Mining in PISA CBA 2012 Problem Solving Dataset Using Association Rules**. 2018 IEEE 12th International Symposium on Applied Computational Intelligence and Informatics (SACI). **Anais...IEEE**, 2018

PERI, A. *et al.* Lo que PISA nos mostró: claroscuros de la participación de Uruguay a lo largo de una década. **RELIEVE-Revista Electrónica de Investigación y Evaluación Educativa**, v. 22, n. 1, 2016.

REIS, R. X. DOS. **A utilização do fenômeno Big Data na Administração Pública - a experiência do PENSA na Prefeitura Municipal do Rio de Janeiro**. Trabalho de conclusão de curso—Rio de Janeiro: Escola de Contas e Gestão, 2015.

ROMERO, C. **Handbook of educational data mining**. Boca Raton: Taylor & Francis Group, 2011.

TAŞTAN, H.; ERDOĞAN, S. Cognitive skills and economic performance: evidence from the recent international student assessment tests. **Eurasian Economic Review**, v. 8, n. 3, p. 417–449, dez. 2018.

TELTEMANN, J.; WINDZIO, M. The impact of marketisation and spatial proximity on reading performance: international results from PISA 2012. **Compare: A Journal of Comparative and International Education**, p. 1–18, 27 jun. 2018.