

Descoberta de Conhecimento sobre a Influência Interdisciplinar no Desempenho Discente do Ensino Técnico de Informática Integrado ao Médio

Thales V. Maciel¹, Mariele de A. Lanes², James G. Ramos¹, Leandro K. Wives²

¹Instituto Federal de Educação, Ciência e Tecnologia Sul-rio-grandense (IFSul)
Campus Bagé – Av. Leonel de Moura Brizola – 96.418-400 – Bagé – RS – Brasil

²Centro Interdisciplinar de Novas Tecnologias na Educação (CINTED)
Universidade Federal do Rio Grande do Sul (UFRGS)
Av. Paulo Gama – 90.040-060 – Porto Alegre – RS – Brasil

thalesmaciel@ifsul.edu.br, infomarielle@gmail.com, wives@inf.ufrgs.br

Abstract. *Considering the need for a better understanding of the influence of student's individual performance in high school integrated technical education subjects on the performance in the discipline of algorithms, this study aimed to objectively substantiate interdisciplinarity strategies. Knowledge discovery in databases was performed through data mining linear regression experiments on a dataset containing indicators of students performance in 11 subjects. Using three distinct methods, obtained results indicated that website construction, portuguese language, biology and mathematics' performance as the most influential on students' individual performance in the algorithms subject.*

Resumo. *Considerando a necessidade de compreensão da influência dos desempenhos individuais de alunos do ensino técnico integrado ao ensino médio nas demais disciplinas sobre a disciplina de algoritmos, objetivou-se fundamentar propostas de interdisciplinaridade de forma objetiva, não apenas subjetivamente. Para tanto, procedeu-se à descoberta de conhecimento em banco de dados, especificamente com tarefas de regressão linear da mineração de dados a partir de conjuntos de dados referentes aos desempenhos individuais de alunos em 11 disciplinas. Com três métodos distintos, obteve-se resultados que indicaram as disciplinas de construção de websites, língua portuguesa, biologia e matemática como de maior influência nos respectivos desempenhos na disciplina de algoritmos.*

1. Introdução

Segundo [Rapkiewicz et al. 2007], a disciplina de algoritmos é frequentemente o primeiro contato dos alunos de cursos da área de computação com tais tecnologias e a disciplina comumente trabalha a solução de problemas através da abstração e sem a utilização de uma linguagem de programação funcional, mas, pseudocódigo, como material didático. Ainda assim, os níveis de reprovação e evasão são altos, o que reflete uma das problemáticas enfrentadas pelas instituições que oferecem cursos na área de computação.

Neste sentido, a importância da interdisciplinaridade no ensino básico está na inovação e práticas de ações coletivas dos docentes no processo educativo, que pode servir como estratégia de combate a evasão [Terradas 2019].

O problema de pesquisa abordado neste estudo é definido em "como o desempenho individual dos alunos nas demais disciplinas do ensino técnico em informática integrado ao ensino médio influencia nos respectivos desempenhos na disciplina de algoritmos?"

Em consonância com a bibliografia de trabalhos correlatos, trabalha-se a hipótese de que as maiores influências ao desempenho dos alunos na disciplina de algoritmos ocorrem pelos seus desempenhos nas disciplinas de matemática e aquelas de natureza técnica como, por exemplo, introdução a informática.

Este trabalho busca fundamentar propostas de interdisciplinaridade com base em dados e análise de forma objetiva, não apenas subjetiva, e que a metodologia descrita possa ser replicada para melhoria dos planos de ensino das demais disciplinas das grades curriculares existentes.

O restante do trabalho está organizado da seguinte forma: na Seção 2 são elencados os trabalhos relacionados a esta pesquisa; a Seção 3 apresenta a metodologia empregada no estudo; na Seção 4 estão descritos os resultados obtidos; e a Seção 5 apresenta as conclusões e os trabalhos futuros.

2. Trabalhos Correlatos

Segundo [Silva et al. 2017] as habilidades acadêmicas fundamentais, tais como: leitura e escrita, interpretação e cálculos e raciocínio matemático devem ser desenvolvidas no período da educação básica dos estudantes. De acordo com [Amaral et al. 2017], os estudantes são incapazes de desenvolver o raciocínio lógico para propor soluções algorítmicas, assim como para abstração necessária na compreensão do problema, o que acaba gerando um grande número de evasões e reprovações nas disciplinas de algoritmos e programação de computadores.

Nesse sentido, é provável que as dificuldades que os estudantes apresentam na aprendizagem dos conteúdos em programação podem estar relacionadas a dificuldades da educação básica. Para investigar a relação existente entre o desempenho obtido pelos estudantes na disciplina de algoritmos e programação com o seu desempenho no ensino médio, [Schorr and Bercht 2018] apresentaram um estudo longitudinal de estudantes do ensino médio frente as suas médias na disciplina de algoritmos e programação nos cursos da área da computação.

Para isso, foi realizada uma análise das médias de cada disciplina do ensino médio e da disciplina de algoritmos e programação de 137 alunos dos cursos de sistemas de informação, engenharia de software, engenharia da computação e redes de computadores, de uma Instituição de Ensino Superior.

Os resultados demonstram que os alunos que, durante o ensino médio, conseguiram melhor desempenho nas disciplinas de matemática, física e português são aqueles que tiveram melhores resultados em algoritmos e programação. Isto demonstra que competências desenvolvidas nestas disciplinas auxiliam na aprendizagem de algoritmos e programação.

Considerando a importância da interdisciplinaridade para a institucionalização de novos modelos de educação e formação, além da gestão interna das organizações e do mercado de trabalho em geral, [da Silva 2017] propôs analisar a possibilidade da inter-

disciplinaridade no ensino aprendizagem dos cursos superiores de tecnologia como um mecanismo facilitador às exigências do mundo do trabalho, visando identificar os níveis e as relações de interação entre as disciplinas do curso superior de tecnologia; ponderar as limitações e as vantagens da prática interdisciplinar dos professores/estudantes e compreender a interação entre as disciplinas do curso superior de tecnologia frente à realidade do mundo do trabalho.

Para isso, foi utilizado o método do estudo de caso com abordagem quantitativa, através da aplicação de questionários com perguntas semiestruturadas junto aos professores e estudantes das disciplinas do Curso Superior de Tecnologia em Sistemas para Internet, do Instituto Federal de Ciência e Tecnologia de Mato Grosso do Sul (IFMS). Através da análise de conteúdo são interpretadas as principais características, percepções e representações dos atores envolvidos sobre os níveis de interdisciplinaridade presentes no curso superior de tecnologia. Entre os resultados deste estudo, destacaram-se as vantagens da interação das disciplinas do curso como o esforço conjunto de seus professores e estudantes.

Como limitação dessas pesquisas, percebe-se que [Schorr and Bercht 2018] poderiam ter utilizado algoritmos de mineração de dados para o cruzamento das notas, proporcionando uma melhor análise dessas informações. [da Silva 2017] analisou apenas dados obtidos através da aplicação de questionários, sem levar em consideração o desempenho individual dos estudantes nas diferentes disciplinas.

3. Metodologia

Para tal, procedeu-se a descoberta de conhecimento em banco de dados (DCBD), que consiste na atividade de transformar grandes quantidades de dados em informação que possa ser utilizada na prática, com relevância e novidade em determinada área de conhecimento [Tan et al. 2005]. Segundo [Maciel et al. 2015], o objetivo da DCBD é extrair um modelo útil, com propósito definido e que possibilite o entendimento humano, como, por exemplo, prever a ocorrência de uma importante situação em determinado domínio de conhecimento.

Segundo [Tan et al. 2005], processo da DCBD é formado por três etapas: pré-processamento, onde os dados são preparados desde seu estado bruto com tarefas de limpeza, integração, redimensionamento e transformação até a aplicação de um algoritmo de processamento; processamento, onde algoritmos de mineração de dados são aplicados sobre o conjunto de dados já pré-processado, com o objetivo da descoberta de um modelo preditivo ou descritivo; e o pós-processamento, onde o modelo descoberto no processamento é avaliado qualitativamente e apresentado em formato compreensível por humanos. A Figura 1 ilustra o fluxo de atividades deste processo.

A mineração de dados é definida por [Witten et al. 2016] como a extração de conhecimento implícito, novo e relevante, pela elaboração de programas de computador que percorrem bases de dados de modo automatizado em ambiente computacional em busca de padrões de dados. Tarefas da mineração de dados incluem, proeminentemente, a descoberta de regras de associação, classificação e indução de árvores de decisão, regressão e análise de agrupamentos, onde são descobertos padrões de itens frequentes, modelos para previsão de categorias, estruturas para previsão de valores numéricos e definição de agrupamentos de dados, respectivamente [Tan et al. 2005].

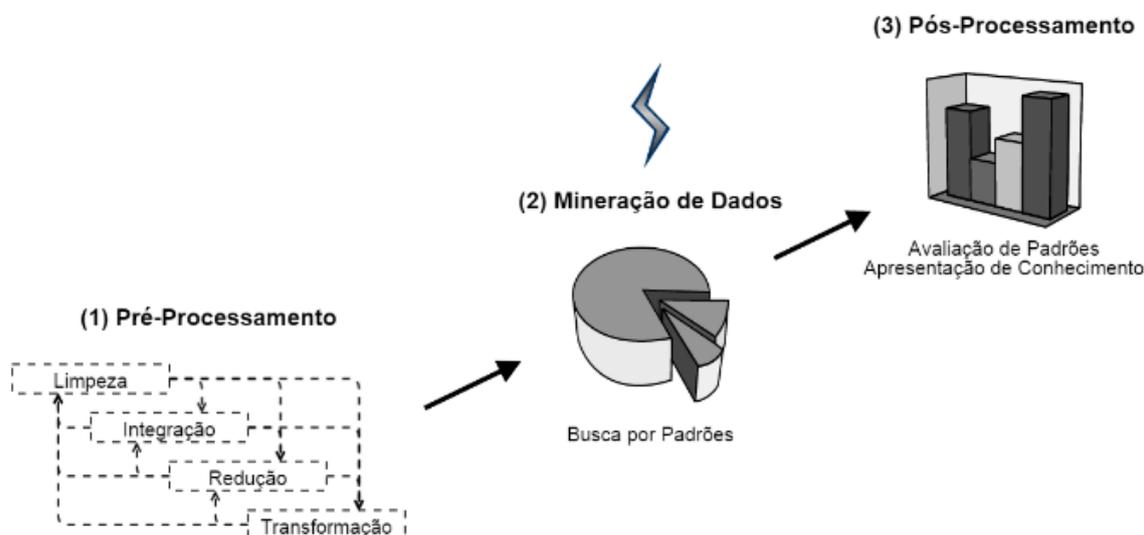


Figura 1. Mineração de dados no contexto do processo de descoberta de conhecimento [Maciel et al. 2015].

A mineração de dados educacionais (*Educational Data Mining*, EDM) é uma área de pesquisa que tem como principal objetivo o desenvolvimento de métodos para explorar conjuntos de dados coletados em ambientes educacionais, visando o melhor entendimento do processo de ensino aprendizagem dos alunos [Baker et al. 2011]. Nos últimos anos a EDM vem se estabelecendo como uma forte e consolidada linha de pesquisa que possui grande potencial para melhorar a qualidade do ensino. Sendo assim, é importante analisar e compreender o que os dados escolares/acadêmicos podem revelar sobre uma instituição de ensino, seu corpo docente e discente.

O conjunto de dados analisado no presente estudo consiste nos conceitos atribuídos nos boletins de desempenho escolares de 117 alunos do primeiro semestre do ensino técnico integrado ao ensino médio de uma instituição federal de ensino técnico do Brasil de 4 períodos distintos (2017 à 2018). As disciplinas cursadas por estes alunos no período foram 11: algoritmos, educação artística, biologia, construção de *websites*, filosofia, física, língua inglesa, língua portuguesa, matemática, metodologia científica e química e os conceitos em formato numérico variam na mesma escala desde 0 (menor avaliação possível) e 10 (maior avaliação possível). Cada instância do conjunto de dados representa um aluno do primeiro semestre no período e cada atributo representa a nota do respectivo aluno em uma disciplina.

Desta forma, projeta-se experimentos para descoberta de modelos preditivos tendo como alvo o conceito numérico na disciplina de algoritmos e sendo atributos preditores os conceitos numéricos referentes a todas as demais disciplinas contidas no conjunto de dados.

No contexto do presente estudo, apreciou-se a tarefa de mineração de dados da regressão. Segundo [Witten et al. 2016], quando o atributo alvo de predição e os preditivos são numéricos, a regressão linear clássica é a técnica de mineração de dados que deve ser naturalmente considerada. Na regressão linear, tem-se o objetivo de expressar o atributo alvo em termos da combinação entre os demais atributos com pesos calculados a partir dos próprios dados fornecidos. Desta forma, além da regressão linear tradicio-

nal, também foram realizados experimentos de regressão linear pelo método de mínimos quadrados médios na etapa de processamento.

Para fins de realização das tarefas e experimentos descritos neste estudo, foi empregado o *Waikato Environment for Knowledge Analysis (WEKA)*, que é um ambiente para análise de conhecimento onde está contida uma coleção de implementações de algoritmos para utilização em atividades de mineração de dados [Witten et al. 2016].

Em pré-processamento, fora identificada a ocorrência de uma instância onde o aluno teria recebido conceito 0 para todas disciplinas. Uma possibilidade desta ocorrência é por reprovação do aluno por infrequência no período ou, mesmo, por evasão. Desta forma, a instância fora classificada como ruído e removida do conjunto de dados antes do processamento, de forma que 116 instâncias permaneceram no conjunto de dados para a atividade de processamento.

Quando da realização da mineração de dados pelo método tradicional de regressão linear, foi obtido um modelo conforme apresentado na Tabela 1. O primeiro experimento teve o algoritmo de processamento configurado para não realizar seleção automatizada de atributos.

Tabela 1. Modelo descoberto pelo primeiro experimento descrito.

Experimento I
Algoritmos =
-0.0334 * Artes +
0.2546 * Biologia +
0.0903 * Filosofia +
0.0544 * Física +
0.3756 * Construção de <i>Websites</i> +
-0.1895 * Língua Inglesa +
0.2777 * Língua Portuguesa +
0.2368 * Matemática +
0.0539 * Metodologia Científica +
-0.0314 * Química +
-0.4332

O modelo apresentado na Tabela 1 mostrou que o desempenho dos alunos na disciplina de construção de *websites* tem maior influência nos desempenhos individuais na disciplina de algoritmos, em comparação com as demais disciplinas ofertadas no semestre. Após esta, língua portuguesa, biologia, matemática, língua inglesa, filosofia, física, metodologia científica, educação artística e, por último, química, foram apresentadas como aquelas cujo desempenho dos alunos têm influência no desempenho em algoritmos, em ordem decrescente sucessivamente.

O segundo experimento também foi realizado pelo método tradicional de regressão linear, mas configurado para realizar a seleção automatizada de atributos, de forma a eliminar da composição do modelo aqueles atributos de menor coeficiente de padronização [Witten et al. 2016]. O modelo resultante deste experimento é apresentado na Tabela 2.

O modelo descoberto (Tabela 2) descreve que, das disciplinas ofertadas no pri-

Tabela 2. Modelo descoberto pelo segundo experimento.

Experimento II
Algoritmos = $0.261 * \text{Biologia} +$ $0.3772 * \text{Construção de } \textit{websites} +$ $-0.1567 * \text{Língua Inglesa} +$ $0.3109 * \text{Língua Portuguesa} +$ $0.2559 * \text{Matemática} +$ -0.1392

meiro semestre do curso técnico de informática integrado com ensino médio, apenas as disciplinas de biologia, construção de *websites*, língua inglesa, língua portuguesa e matemática apresentam relação com a disciplina de algoritmos no tocante ao desempenho dos alunos. Em termos quantitativos, o modelo apresenta que a maior influência no desempenho na disciplina de algoritmos está na disciplina de construção de *websites*, seguido de língua portuguesa, biologia, matemática e língua inglesa, sucessivamente.

O terceiro experimento foi realizado pelo método de regressão linear por mínimos quadrados médios e o modelo descoberto é apresentado na Tabela 3.

Tabela 3. Modelo descoberto pelo terceiro experimento.

Experimento III
Algoritmos = $-0.1016 * \text{Artes} +$ $0.4851 * \text{Biologia} +$ $0.1973 * \text{Filosofia} +$ $0.0228 * \text{Física} +$ $0.1797 * \text{Construção de } \textit{Websites} +$ $-0.0389 * \text{Língua Inglesa} +$ $0.0625 * \text{Língua Portuguesa} +$ $0.2278 * \text{Matemática} +$ $0.1209 * \text{Metodologia Científica} +$ $-0.0292 * \text{Química} +$ -0.6841

O modelo descoberto (Tabela 3) apresentou a influência do desempenho em todas as demais disciplinas no desempenho da disciplina de algoritmo, o que é característica do método utilizado. Segundo o modelo, o desempenho de maior influência no desempenho na disciplina de algoritmos está na disciplina de biologia, seguido de matemática, filosofia, construção de *websites*, metodologia científica, educação artística, língua portuguesa, língua inglesa, química e, por último, física, sucessivamente.

4. Resultados Obtidos

Os experimentos descritos na metodologia tiveram seus resultados avaliados com o método de testes por validação cruzada em 6 partições. O modelo descoberto a partir da

regressão linear tradicional (RLT) apresentou coeficiente de correlação e erro médio absoluto de 0,7805 e 1,0608, respectivamente, no primeiro experimento e 0,7863 e 1,0578, respectivamente, quando configurado para realizar a seleção automatizada de atributos (RL-SA) para o modelo. O modelo descoberto a partir da regressão linear pelo método de mínimos quadrados médios (RLMQM) apresentou coeficiente de correlação de 0,7851 e erro médio absoluto de 1,0905, indicando que este foi um modelo de qualidade ligeiramente inferior ao anterior. A Tabela 4 apresenta estes dados para comparação visual.

Tabela 4. Comparação de indicadores de qualidade dos modelos descobertos.

	Coef. Corr.	Erro Méd. Ab.
RLT	0,7805	1,0608
RL-SA	0,7863	1,0578
RLMQM	0,7851	1,0905

Acerca do modelos descobertos nos experimentos primeiro e segundo, entende-se que o desempenho na disciplina de construção de *websites* ser o de maior influência no desempenho em algoritmos se deve a que são as únicas duas disciplinas técnicas da informática no semestre e que por tal motivo desperta o interesse concomitante dos alunos, justamente por estarem iniciando uma formação técnica.

O desempenho na disciplina de língua portuguesa aparece como de segunda maior relevância nos modelos descobertos nos experimentos primeiro e segundo para previsão do desempenho em algoritmos e entende-se que isto se deve ao método de ensino praticado nas disciplinas. Enquanto a disciplina de algoritmos trabalha a construção de lógica por meio de construtores do português estruturado [Manzano 2010], a disciplina de língua portuguesa visa a interpretação de texto e os próprios construtores da língua portuguesa.

Ao analisar a ementa da disciplina de biologia, é possível encontrar conteúdos que exercitam o raciocínio lógico dos alunos, como os cálculos relacionados aos genes alelos recessivos e dominantes. Em [Ferreira et al. 2015], os autores propõem a analogia entre as definições dos processos biológicos diversos e a especificação de uma sequência ordenada de tarefas para se resolver determinado problema, que vem a ser, justamente, um conceito fundamental para a compreensão de algoritmos. Desta forma, relaciona-se a maior facilidade no desenvolvimento das competências abordadas na disciplina de algoritmos ao estudo destes conteúdos na disciplina de biologia. Esta aparece como de maior influência de acordo com o terceiro modelo descoberto no estudo.

A disciplina de matemática tem o desempenho descrito como influente no desempenho na disciplina de algoritmos nos três modelos descobertos e tem-se que os exercícios de cálculos matemáticos podem ser explorados em interdisciplinaridade com algoritmos, visto que podem ser desenvolvidos programas para automatização dos mesmos, a título de exercício didático.

5. Conclusão e Trabalhos Futuros

Este trabalho objetivou fundamentar propostas de interdisciplinaridade com base em dados e análise de forma objetiva, especificamente com um estudo de caso sobre a disciplina de algoritmos no contexto do ensino técnico integrado ao ensino médio. Utilizando dados

oriundos de uma instituição da rede federal de ensino técnico, foi proposta uma abordagem baseada em descoberta de conhecimento em banco de dados com aplicações da mineração de dados para descobrir a influência do desempenho dos alunos nas demais disciplinas da grade curricular nos respectivos desempenhos na disciplina de algoritmos.

Os resultados obtidos nos experimentos evidenciaram que as disciplinas de construção de *websites*, língua portuguesa e matemática são as quais cujo desempenho individual dos alunos têm maior influência nos respectivos desempenhos na disciplina de algoritmos, respondendo, portanto, o problema de pesquisa apresentado. A disciplina de biologia ainda foi apresentada com maior proeminência no modelo descoberto pelo terceiro experimento realizado, em comparação com os modelos descobertos pelos experimentos primeiro e segundo.

Tendo identificado influências individuais dos desempenhos das demais disciplinas nos respectivos desempenhos na disciplina de algoritmos, considera-se o objetivo da pesquisa cumprido, visto que esta relação é evidenciada de forma objetiva, provendo precedente para estratégias de interdisciplinaridade entre os componentes curriculares dos cursos. Isto é evidenciado pelo formato didático e intuitivo dos modelos descobertos no estudo.

Os resultados obtidos também corroboram com a hipótese apresentada, onde é sugerido que os desempenhos em matemática e língua portuguesa influenciam no desempenho em algoritmos e propõe que outras disciplinas também apresentam este tipo de influência.

Trabalhos futuros envolvem o aumento de conjunto de dados e aplicação da metodologia proposta para descoberta de potenciais interdisciplinaridades entre os demais componentes curriculares do curso e aplicação em outros níveis de ensino.

Referências

- Amaral, E., Camargo, A., Gomes, M., Richa, C. H., and Becker, L. (2017). Algo+ uma ferramenta para o apoio ao ensino de algoritmos e programação para alunos iniciantes. In *Brazilian Symposium on Computers in Education (Simpósio Brasileiro de Informática na Educação-SBIE)*, volume 28, page 1677.
- Baker, R., Isotani, S., and Carvalho, A. (2011). Mineração de dados educacionais: Oportunidades para o Brasil. *Brazilian Journal of Computers in Education*, 19(02):11.
- da Silva, M. O. (2017). A interdisciplinaridade como uma possibilidade no processo ensino aprendizagem da educação profissional de nível tecnológico para o mundo do trabalho. *Revista Brasileira da Educação Profissional e Tecnológica*, 2(13):13–30.
- Ferreira, A. C., Melhor, A., Barreto, J., de Paiva, L. F., and Matos, E. (2015). Experiência prática interdisciplinar do raciocínio computacional em atividades de computação desplugada na educação básica. In *Anais do Workshop de Informática na Escola*, volume 21, page 256.
- Maciel, T. V., da Rosa Seus, V., dos Santos Machado, K., and Borges, E. N. (2015). Mineração de dados em triagem de risco de saúde. *Revista Brasileira de Computação Aplicada*, 7(2):26–40.

- Manzano, J. A. N. (2010). *Algoritmos lógica para desenvolvimento de programação de computadores*. Saraiva Educação SA.
- Rapkiewicz, C. E., Falkembach, G. A. M., Seixas, L. M. J. d., Santos, N. d. S. R. S., Cunha, V. V. d., Klemann, M., et al. (2007). Estratégias pedagógicas no ensino de algoritmos e programação associadas ao uso de jogos educacionais. *RENOTE: revista novas tecnologias na educação [recurso eletrônico]*. Porto Alegre, RS.
- Schorr, M. and Bercht, M. (2018). Análise longitudinal do desempenho dos estudantes de ensino médio e estudantes de nível superior para algoritmos e programação. In *Anais dos Workshops do Congresso Brasileiro de Informática na Educação*, volume 7, page 550.
- Silva, N. C., Lima, A. C., de Souza, N., and de Sousa, D. F. (2017). Raciocínio lógico nas escolas: Uma introdução ao ensino de algoritmos de programação. In *Anais dos Workshops do Congresso Brasileiro de Informática na Educação*, volume 6, page 1011.
- Tan, P.-N., Steinbach, M., and Kumar, V. (2005). *Introduction to Data Mining*. Addison-Wesley.
- Terradas, R. D. (2019). A importância da interdisciplinaridade na educação matemática. *Revista da Faculdade de Educação*, 14(16):95–114.
- Witten, I. H., Frank, E., Hall, M. A., and Pal, C. J. (2016). *Data Mining: Practical machine learning tools and techniques*. Morgan Kaufmann.