

***Deep Learning* para Previsão do Desempenho do Estudante: Um Mapeamento Sistemático da Literatura**

Nilcimar Neitzel Will¹, Avaniilde Kemczinski^{1,2}, Rafael Stubs Parpinelli¹

¹ Prog. de Pós-graduação em Computação Aplicada – PPGCA
Universidade do Estado de Santa Catarina (UDESC) - Joinville, SC - Brasil

²Prog. de Pós-graduação em Ensino de Ciências Matemática e Tecnologias - PPGECMT
Universidade do Estado de Santa Catarina (UDESC) - Joinville, SC - Brasil

nil_cc@yahoo.com.br, {avanilde.kemczinski,rafael.parpinelli}@udesc.br

Abstract. *It is not of today that school dropout has been a problem, both in pre-sential environments and in virtual learning environments. Affecting not only the career of the students but also the educational institutions. Predicting student performance helps institutions identify the factors that lead to evasion, and to take measures for student retention at school. This work seeks to study how Deep Learning has been applied in predicting student performance, and to identify how it has been implemented, techniques, methods, algorithms, and tools.*

Resumo. *Não é de hoje que a evasão escolar tem sido um problema, tanto em ambientes presenciais, quanto em ambientes virtuais de aprendizagem. Afetando não apenas a carreira dos estudantes, como também as instituições de ensino. Prever o desempenho do aluno ajuda as instituições a identificar os fatores que levam a evasão, e assim tomar medidas para a retenção do aluno na escola. Este trabalho busca estudar como o Deep Learning tem sido aplicado na previsão do desempenho do estudante, e identificar como ele tem sido implementado, as técnicas, métodos, algoritmos e ferramentas.*

1. Introdução

Segundo Devasia et al. (2016), a retenção dos alunos em uma instituição de ensino está diretamente relacionada ao seu desempenho acadêmico. Identificar os fatores que afetam o desempenho do estudante ajudam as instituições a oferecerem uma educação personalizada aos alunos de risco, como também o oferecimento de bolsas para alunos que se destacam. Assim é possível melhorar a eficiência do sistema e reduzir despesas do processo educacional. Porém, segundo Guo et al. (2015) medir o desempenho do estudante é um desafio, uma vez que depende de diversos fatores, como demografia, antecedentes pessoais, educacionais, psicológicos, progresso acadêmico, frequência do aluno nas aulas até a renda familiar e a qualificação da mãe. Por isso, tal previsão também é conhecida como o “problema dos mil fatores”, portanto os dados coletados neste campo geralmente têm alta dimensionalidade [Nagy e Molontay 2018].

A área que analisa estes dados para fazer previsões sobre o futuro chama-se Análise Preditiva [Patil et al. 2017]. Já a área que se preocupa em desenvolver métodos para lidar com estes dados a partir de ambientes educacionais é conhecida como *Educational Data Mining* (EDM) [Romero e Ventura 2013], ambas incorporam técnicas de *Data*

Mining, Aprendizado de Máquina e Estatística para explorar tais dados e detectar padrões ocultos que caracterizam o comportamento e o desempenho dos alunos [Martins et al. 2017]. Diversas técnicas em *Data Mining* têm sido usadas para previsão de desempenho, entre as principais pode-se citar: *Decision Tree*, *Naive Bayes*, KNN (*K-Nearest Neighbor*), modelos de regressão e Redes Neurais Artificiais. Para Guo et al. (2015), técnicas tradicionais de *Data Mining* podem não ser as mais indicadas ao problema, devido ao fato da não linearidade entre os fatores que envolvem a predição do desempenho acadêmico. Nos últimos anos a tendência de *Deep Learning* (DL) vem ganhando força, especialmente pela capacidade de identificar padrões em bases complexas. Segundo Coelho e Silveira (2017), DL tem mostrado grande potencial em várias tarefas de *Data Mining* relacionadas a classificação, regressão e previsão, ampliando a aplicação das Redes Neurais Artificiais (RNAs), amplamente utilizadas na área de EDM e *Learning Analytics* (LA).

Nesse sentido, Coelho e Silveira (2017) realizam uma revisão de literatura sobre a aplicação de DL à EDM e LA, para saber se as técnicas de DL têm sido usadas e quais as principais questões de pesquisa que o DL abordou. Os autores analisaram 45 trabalhos, sendo 6 de DL e destes, apenas 3 para previsão do desempenho do estudante. Eles destacam o poder alcançado por DL mas concluem que ainda é cedo para julgar o seu impacto na pesquisa em LA e EDM, visto a pequena produção na área de DL até o momento e a dificuldade de análise dos resultados.

Por outro lado, Wilson et al. (2016) faz uma análise de 3 trabalhos, com o propósito de comparar as técnicas de *Deep Knowledge Tracing* (DKT) e *Bayesian Knowledge Tracing* (BKT), e verificar se *Deep Learning* é realmente a "panaceia" para previsão de desempenho de estudante. Sendo BKT um modelo para rastreamento de conhecimento Bayesiano, utilizado há mais de 20 anos, e DKT um modelo alternativo para o BKT baseado em aprendizagem profunda, que utiliza arquitetura de rede recorrente LSTM (*Long Short Term Memory*). Em sua pesquisa, os autores identificaram alguns problemas que, quando não abordados adequadamente, produzem resultados que favorecem o DKT, o que justificaria seu alto desempenho comparado ao BKT. Mesmo sem considerar estes casos, os autores também concluem que os ganhos com o DKT não compensam as dificuldades em interpretar os resultados do modelo.

Ambos os trabalhos citados não foram muito convincentes quanto ao uso de DL para resolver problemas de educação, um diz que este não é a melhor solução, o outro encontrou pouquíssimos resultados. Dessa forma, propõem-se neste trabalho, um mapeamento sistemático da literatura para identificar se mais trabalhos tem surgido na área de DL para previsão de desempenho do estudante, bem como sua implementação, as técnicas, métodos, algoritmos e ferramentas, além das dificuldades encontradas e se houve avanços com relação a interpretação dos resultados. Para contextualizar, na Seção 2 é dada a definição de DL. Na Seção 3, é descrita a metodologia de pesquisa adotada para este trabalho. Em seguida, os resultados obtidos são relacionados e discutidos na Seção 4. Por fim, as considerações finais na Seção 5, seguida dos agradecimentos e das referências bibliográficas.

2. *Deep Learning*

Deep Learning (DL), ou Aprendizagem Profunda, pode ser definido como uma subárea do Aprendizado de Máquina, caracterizado pela utilização de várias camadas de informações

para extração de características, transformação e análise de padrões. Também se caracteriza pelo uso de algoritmos de aprendizado que buscam identificar relações entre os dados através de modelos estatísticos [Jost 2015].

Segundo Xing e Du (2019), o DL processa os dados de entrada através de um grande número de camadas ocultas em sua estrutura. Cada camada é feita de nós, onde ocorrem os cálculos. Um conjunto de pesos é combinado às entradas, para atribuir-lhes significância. Esses pesos de entrada são somados e avaliados para decidir que medida a informação se propaga através da rede. Ou seja, a camada oculta adapta o conjunto de recursos usando a saída da camada anterior. Quanto mais camadas ocultas tiver, mais complexos e abstratos serão os dados.

Para Xing e Du (2019), o DL pode alcançar melhor desempenho de previsão comparado a outros algoritmos, o que pode ser refletido tanto na precisão da previsão quanto na personalização da intervenção. Além disso o DL pode ser aplicado no reconhecimento de voz, no processamento de linguagem natural, visão computacional e rastreamento de conhecimento. Este último, é um dos principais componentes para permitir o aprendizado e o ensino personalizado, onde os alunos podem escolher o que precisam para praticar mais e os professores podem decidir quais conceitos ensinar mais aos alunos [Chen et al. 2018].

3. Metodologia

Foi realizado um mapeamento sistemático seguindo a metodologia proposta por Petersen et al. (2008), que é composta por 4 etapas descritas na sequência.

3.1. Questões de Pesquisa

Através deste mapeamento procura-se responder à pergunta "Como DL tem sido usado na previsão de desempenho do estudante?". Para auxiliar a responder esta pergunta, foram definidas as seguintes questões de pesquisa:

- Quais os objetivos dos trabalhos?
- Quais as técnicas, métodos, algoritmos e ferramentas utilizadas?
- Quais as dificuldades encontradas na execução dos trabalhos?
- Quais os resultados alcançados?

Na próxima etapa é detalhado como foi realizada a busca a fim de encontrar trabalhos que respondam estas questões.

3.2. Condução da Busca

A busca por trabalhos foi conduzida em 4 Mecanismos de Busca Acadêmica (MBAs), dentre os mais bem avaliados segundo Buchinger et al. (2014), são eles: IEEE, ACM, Science Direct e Scopus. A pesquisa foi realizada nesta ordem.

Foram realizados diversos testes nos mecanismos citados, até chegar numa string de busca que obtivesse os resultados mais adequados segundo as questões de pesquisa. A cada teste realizado era feita uma análise nos primeiros 10 resultados, para verificar se título, resumo e palavras-chave estavam de acordo.

Dessa forma, a busca foi realizada através de título, resumo e palavras-chave em cada mecanismo, utilizando a seguinte string: (((“deep learning”) AND (technique OR

method OR approach OR application OR tool OR analysis)) AND (predict* AND (student* OR academic))). O asterisco (*) é um caractere coringa que dependendo o mecanismo deve ser substituído por interrogação (?).

Na próxima etapa são abordados os critérios de inclusão e exclusão que foram aplicados nos resultados encontrados.

3.3. Seleção dos Trabalhos

Os critérios de inclusão (CI) e os critérios de exclusão (CE) definidos são apresentados na Tabela 1 :

Tabela 1. Critérios de inclusão e exclusão

CI	CE
1. Escritos em inglês, português ou espanhol;	1. Artigo duplicado;
2. Artigos completos com 4 ou mais páginas;	2. Trabalho duplicado: artigos que tratam do mesmo trabalho;
3. Artigo disponível para download;	3. Trabalho sem relação com a proposta.
4. Estudo primário.	

De todos os artigos resultantes foram lidos o título, resumo e palavras-chave e aplicados os critérios de inclusão e exclusão. Os resultados são apresentados na Tabela 2:

Tabela 2. Resultados da busca após aplicação dos critérios

Mecanismo	Resultados da busca	Após CI e CE
IEEE	56	12
ACM	23	9
Science Direct	16	1
Scopus	134	3
Total	228	25

Na Tabela 2, o mecanismo Scopus apresenta o maior número de resultados antes da aplicação dos critérios, porém o menor número após aplicação dos critérios. Isso se deve ao fato de que este mecanismo retorna como resultado diversos artigos dos mecanismos analisados anteriormente. Então, ao aplicar os critérios de exclusão, todos os trabalhos repetidos nos outros mecanismos foram descartados, resultando apenas 3 trabalhos diferentes.

Todos os trabalhos analisados encontram-se referenciados no link: https://drive.google.com/file/d/1zGoyxkTHTjJvPw_DNE01rb5v1VYrpdsn/view?usp=sharing.

3.4. Extração e Mapeamento dos Dados

Todos os trabalhos selecionados foram estudados e, para responder as questões de pesquisa foram extraídos os seguintes dados: objetivo; métricas de avaliação; arquitetura de rede; ferramentas e resultados. Na próxima seção é apresentada a análise dos resultados.

4. Resultados e Análises

Primeiramente, foram extraídos os dados bibliométricos com relação ao ano de publicação dos artigos. Na Figura 1, pode-se observar que há um crescimento na produção anual de artigos relacionados a previsão de desempenho de estudante utilizando DL. Considerando que esta pesquisa foi realizada em meados de 2019, até o final deste ano o número de artigos deve ultrapassar o do ano anterior, caso a tendência observada se confirme. Com relação a Coelho e Silveira (2017), onde foram encontrados apenas 6 artigos de DL, pode-se dizer que houve um aumento de mais de 300% desde 2017 até o momento.

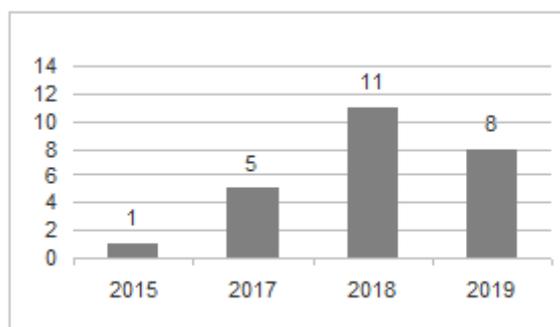


Figura 1. Publicação por ano

Com relação aos objetivos dos trabalhos analisados, foram encontrados 3 objetivos distintos, que são apresentados na Figura 2. Cada trabalho está relacionado a apenas um objetivo no gráfico. O objetivo principal é propor um modelo/abordagem/sistema. É interessante destacar que no trabalho de Hu e Rangwala (2019) são propostos 2 modelos, um com base em MLP (*Multilayer Perceptron*) e outro em LSTM, a ideia é associar a estimativa de incerteza às previsões, e a influência dos cursos anteriores na previsão do curso atual. Já em Xing e Du (2019) os autores têm 2 objetivos, explorar o uso das técnicas de DL e investigar a intervenção personalizada através do uso de DL.

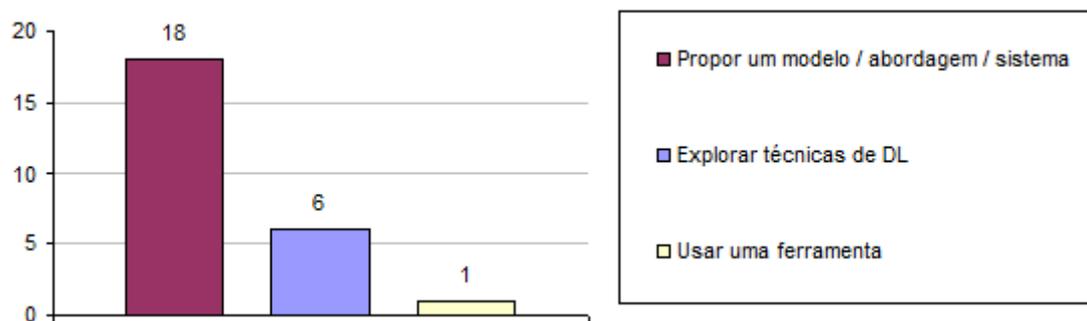


Figura 2. Objetivos dos trabalhos

Com relação as técnicas, métodos, algoritmos e ferramentas utilizadas nos trabalhos analisados, neste trabalho optou-se por extrair os dados relacionados as métricas de avaliação, a arquitetura de rede utilizada para implementar o DL e as ferramentas que foram utilizadas para o desenvolvimento destas redes.

As métricas de avaliação servem para medir a qualidade do modelo de acordo com os objetivos especificados, bem como ajudam a avaliar a capacidade de erro e acerto

dos modelos. Na Figura 3, são apresentadas as métricas de avaliação identificadas nos trabalhos analisados. Ao todo foram identificadas 9 métricas. A maioria dos trabalhos utilizaram mais de uma métrica, uma vez que não se recomenda a utilização de uma única medida para avaliação dos modelos. Sendo *Precision* e AUC as mais utilizadas. *Precision*, ou valor preditivo positivo, é a fração de instâncias relevantes entre as instâncias recuperadas [Hussain et al. 2019]. AUC é um método gráfico com aplicação em diversas áreas, como medicina, economia, e na Aprendizagem de Máquina para avaliação de modelos de classificação.

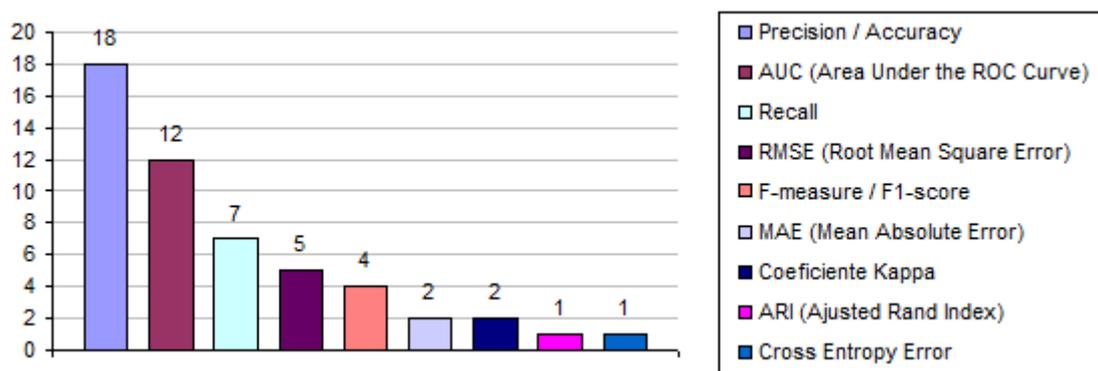


Figura 3. Métricas de avaliação identificadas nos trabalhos

Na Figura 4, são apresentadas as arquiteturas de rede utilizadas nos trabalhos para a implementação do DL. Foram identificados 6 tipos de arquitetura de rede. Cada trabalho corresponde a apenas uma arquitetura no gráfico. As arquiteturas mais utilizadas foram a *Feed Forward* e a LSTM. Sendo a rede LSTM um tipo de rede recorrente, tem a característica de resolver problemas de dependência a longo prazo, uma vez que lembrar das informações por longos períodos é seu comportamento padrão, diferente das redes tradicionais como *Feed Forward* [Anantharaman et al. 2018]. Em Wang et al. (2017), é proposto um modelo DL que combina CNN (Redes Neurais Convolucionais) e RNN (Redes Neurais Recorrentes) para extrair automaticamente os dados brutos de MOOCs (*Massive Open Online Course*), para previsão de abandono.

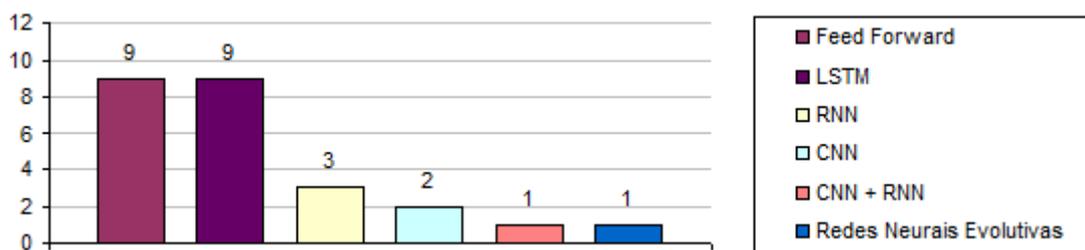


Figura 4. Arquiteturas de rede identificadas nos trabalhos

Na Figura 5, são apresentadas as ferramentas que foram utilizadas nos trabalhos para o desenvolvimento das redes. Apenas 8 trabalhos mencionaram as ferramentas utilizadas. Alguns trabalhos utilizaram mais de uma ferramenta. Na imagem, pode-se observar que a biblioteca Google TensorFlow é a mais utilizada, somando 6 trabalhos. Segundo Fok et al. (2018) o TensorFlow tem ganhado popularidade por ser uma biblioteca de

código aberto para Aprendizado de Máquina, mas geral o suficiente para uma diversidade de outras aplicações, além de ser multiplataforma e poder ser utilizado desde pesquisas a grandes empresas.

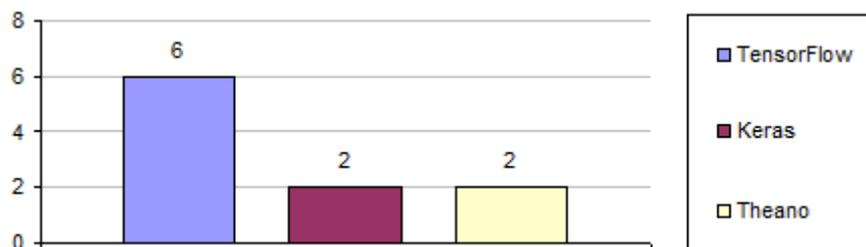


Figura 5. Ferramentas identificadas nos trabalhos

Na Figura 6 é apresentada uma classificação das bases de dados utilizadas nos trabalhos revisados nas categorias Ensino à Distância (EAD), Presencial e Outros. A maioria dos trabalhos utilizou base de dados de cursos EAD, como o ShareCourse que é uma plataforma MOOC chinesa que fornece análise de dados e ferramentas de visualização para auxiliar o aprendizado [Huang et al. 2018]. Na coluna Presencial se encaixam os trabalhos que utilizaram dados de sistemas acadêmicos e escolares de cursos presenciais, como informações dos alunos e notas. Em Outros, se encaixam os trabalhos que utilizaram dados de sistemas que não estão diretamente ligados a uma universidade ou escola, como por exemplo dados de vestibular do SISU ou ENEM.



Figura 6. Bases de dados identificadas nos trabalhos

Com relação aos resultados dos trabalhos, todos relataram um bom desempenho dos modelos DL com relação as métricas utilizadas. Com relação a previsão, 18 trabalhos afirmaram que o modelo DL utilizado obteve desempenho superior aos modelos tradicionais.

Sobre os problemas encontrados, dois trabalhos relatam o problema de dados ausentes e inconsistentes ao se trabalhar com dados de *e-learning*. Em Nagy e Molontay (2018) o problema foi resolvido com o algoritmo K-NN para imputação de dados, e o algoritmo de seleção de recursos (FS) para reduzir a redundância. Já em Fouki et al. (2019) foi criada uma etapa de pré-processamento e usado o algoritmo PCA (*Principal Component Analysis*) para compactar o conjunto de dados e preservar as informações. Fouki et al. (2019) também comentam a dificuldade em definir o número de camadas ocultas da rede neural, pois um número muito grande aumenta o tempo de treinamento e retarda o

processo de aprendizado, e um número muito pequeno diminui os recursos de processamento da rede. Para os autores, existem duas formas para escolher o número de camadas: começando com uma rede pequena que vai aumentando, ou com uma rede com várias camadas que vai diminuindo a dimensão do modelo até chegar num modelo ideal.

Com relação a interpretação dos resultados do DL, a maioria dos trabalhos não comentou nada a respeito. Tanuar et al. (2018) destacam que *Decision Tree* ainda é o modelo mais fácil para apresentar os resultados ao usuário. Botelho et al. (2019) e Lee e Yeung (2019) relatam que apesar dos resultados serem menos interpretáveis e do aumento da complexidade, o uso de DL compensa pelas vantagens mencionadas e os resultados mais precisos. Hu e Rangwala (2019) foi o único trabalho que propôs uma abordagem para interpretabilidade dos resultados.

5. Considerações Finais

Com este trabalho pode-se concluir que o DL tem sido usado, e cada vez mais, para previsão de desempenho de estudante. A Figura 1 mostra que sua utilização tem crescido consideravelmente desde 2015. Este estudo também contribui para mostrar as métricas de avaliação utilizadas e a importância da utilização de mais de uma métrica para avaliar o modelo, pois apesar da precisão ser a mais utilizada, nem sempre é a mais adequada. Por isso, a utilização de outras métricas como AUC, Recall e RMSE complementam a validação do modelo.

Este trabalho também mostra que as redes mais utilizadas foram a *Feed Forward* e LSTM, sendo esta última uma boa opção para resolver os problemas de dependência a longo prazo, por utilizar células de memória. Apesar da maioria dos trabalhos não relatarem as ferramentas utilizadas para o desenvolvimento do DL, pode-se concluir que o Google TensorFlow tem se mostrado uma boa opção para implementação de aprendizado de máquina, além de ser uma biblioteca de código aberto.

Para concluir, os resultados dos trabalhos mostram que o DL pode ser sim uma boa opção para resolver os problemas de previsão de desempenho de estudante, e que apesar da complexidade de implementação e dificuldades de interpretação, a precisão dos resultados compensa as dificuldades.

6. Agradecimentos

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Código de Financiamento 001.

Referências

- Anantharaman, H., Mubarak, A. e Shobana, B. T. (2018). Modelling an adaptive e-learning system using lstm and random forest classification. Em *2018 IEEE Conference on e-Learning, e-Management and e-Services (IC3e)*, páginas 29–34.
- Botelho, A. F., Varatharaj, A., Patikorn, T., Doherty, D., Adjei, S. A. e Beck, J. E. (2019). Developing early detectors of student attrition and wheel spinning using deep learning. *IEEE Transactions on Learning Technologies*, 12(2):158–170.
- Buchinger, D., Cavalcanti, G. e Hounsell, M. (2014). Mecanismos de busca acadêmica: uma análise quantitativa. *Revista Brasileira de Computação Aplicada*, 6(1):108–120.

- Chen, P., Lu, Y., Zheng, V. W. e Pian, Y. (2018). Prerequisite-driven deep knowledge tracing. Em *2018 IEEE International Conference on Data Mining (ICDM)*, páginas 39–48.
- Coelho, O. B. e Silveira, I. F. (2017). Deep learning applied to learning analytics and educational data mining: A systematic literature review. Em *VI Congresso Brasileiro de Informática na Educação (CBIE)*, página 143.
- Devasia, T., Vinushree T P e Hegde, V. (2016). Prediction of students performance using educational data mining. Em *2016 International Conference on Data Mining and Advanced Computing (SAPIENCE)*, páginas 91–95.
- Fok, W. W. T., He, Y. S., Yeung, H. H. A., Law, K. Y., Cheung, K., Ai, Y. e Ho, P. (2018). Prediction model for students' future development by deep learning and tensorflow artificial intelligence engine. Em *2018 4th International Conference on Information Management (ICIM)*, páginas 103–106.
- Fouki, M. E., Aknin, N. e Kadiri, K. E. E. (2019). Multidimensional approach based on deep learning to improve the prediction performance of dnn models. *International Journal of Emerging Technologies in Learning (iJET)*, 14(02):30–41.
- Guo, B., Zhang, R., Xu, G., Shi, C. e Yang, L. (2015). Predicting students performance in educational data mining. Em *2015 International Symposium on Educational Technology (ISET)*, páginas 125–128.
- Hu, Q. e Rangwala, H. (2019). Reliable deep grade prediction with uncertainty estimation. Em *Proceedings of the 9th International Conference on Learning Analytics & Knowledge, LAK19*, páginas 76–85, New York, NY, USA. ACM.
- Huang, N., Chen, C., Tzeng, J., Fang, T. T. e Lee, C. (2018). Concept assessment system integrated with a knowledge map using deep learning. Em *2018 Learning With MOOCS (LWMOOCS)*, páginas 113–116.
- Hussain, S., Muhsion, Z., Salal, Y., Theodoru, P., Kurtoğlu, F. e Hazarika, G. C. (2019). Prediction model on student performance based on internal assessment using deep learning. *International Journal of Emerging Technologies in Learning (iJET)*, 14(08):4–22.
- Jost, I. (2015). Aplicação de deep learning em dados refinados para mineração de opiniões. Dissertação de Mestrado, Universidade do Vale do Rio dos Sinos, São Leopoldo.
- Lee, J. e Yeung, D.-Y. (2019). Knowledge query network for knowledge tracing: How knowledge interacts with skills. Em *Proceedings of the 9th International Conference on Learning Analytics & Knowledge, LAK19*, páginas 491–500, New York, NY, USA. ACM.
- Martins, L. C. B., Carvalho, R. N., Carvalho, R. S., Victorino, M. C. e Holanda, M. (2017). Early prediction of college attrition using data mining. Em *2017 16th IEEE International Conference on Machine Learning and Applications (ICMLA)*, páginas 1075–1078.
- Nagy, M. e Molontay, R. (2018). Predicting dropout in higher education based on secondary school performance. Em *2018 IEEE 22nd International Conference on Intelligent Engineering Systems (INES)*, páginas 000389–000394.
- Patil, A. P., Ganesan, K. e Kanavalli, A. (2017). Effective deep learning model to predict student grade point averages. Em *2017 IEEE International Conference on Computational Intelligence and Computing Research (ICCIC)*, páginas 1–6.
- Petersen, K., Feldt, R., Mujtaba, S. e Mattsson, M. (2008). Systematic mapping studies in software engineering. Em *Proceedings of the 12th International Conference on Eva-*

luation and Assessment in Software Engineering, EASE'08, páginas 68–77, Swindon, UK. BCS Learning & Development Ltd.

Romero, C. e Ventura, S. (2013). Data mining in education. *Wiley Int. Rev. Data Min. and Knowl. Disc.*, 3(1):12–27.

Tanuar, E., Heryadi, Y., Lukas, Abbas, B. S. e Gaol, F. L. (2018). Using machine learning techniques to earlier predict student's performance. Em *2018 Indonesian Association for Pattern Recognition International Conference (INAPR)*, páginas 85–89.

Wang, W., Yu, H. e Miao, C. (2017). Deep model for dropout prediction in moocs. Em *Proceedings of the 2Nd International Conference on Crowd Science and Engineering, ICCSE'17*, páginas 26–32, New York, NY, USA. ACM.

Wilson, K. H., Xiong, X., Khajah, M., Lindsey, R. V., Zhao, S., Karklin, Y., Inwegen, E. G. V., Han, B., Ekanadham, C. e Beck, J. E. (2016). Estimating student proficiency: Deep learning is not the panacea. In *Neural Information Processing Systems, Workshop on Machine Learning for Education*.

Xing, W. e Du, D. (2019). Dropout prediction in moocs: Using deep learning for personalized intervention. *Journal of Educational Computing Research*, páginas 547–570.