

## Identificação da Proposta

Título do Projeto:	Educação 4.0 e a Ciência de Dados Educacionais: Análise de desempenho acadêmico baseado em dados inteligentes
Referência da Chamada:	Projeto de Pesquisa para Pós-Doutorado
Linha(s) de atuação em que se insere o projeto:	Ciência de Dados Educacional, Educação 4.0, Aprendizado de Máquina, Mineração de Dados
Candidato:	Renata Pitta Barros
Supervisor do Projeto:	Luiz Marcos Garcia Gonçalves
Instituição Executora:	Universidade Federal do Rio Grande do Norte

## Projeto de Pesquisa

## 1. Título

**Educação 4.0 e a Ciência de dados educacionais: Análise de desempenho acadêmico baseado em dados inteligentes**

## 2. Resumo

A evolução exponencial das tecnologias educacionais tem provocado o surgimento de soluções educacionais que não eram pensadas há alguns anos atrás. Entre as áreas emergentes destaca-se a aprendizagem por dispositivos eletrônicos, mais conhecida na modalidade de Educação à Distância. Em um futuro muito breve a maioria dos dispositivos que nos cercam estarão conectados na Internet, fornecendo informações das mais diversas áreas do conhecimento. Esse cenário irá provocar o surgimento de uma quantidade prodigiosa de dados. Esses por sua vez, precisarão ser processados para gerar informação útil. Nesse contexto, surge uma nova e promissora área de pesquisa, a Ciência dos Dados Educacionais. Uma área que irá impactar diretamente os governos, as escolas, as universidades, as empresas e principalmente, as pessoas. Dessa forma, a formação de recursos humanos em Ciência dos dados educacionais torna-se um ponto estratégico para o desenvolvimento da Educação no país. Nesse sentido, o respectivo projeto propõe a utilização de técnicas da área de Ciência de Dados Educacionais para implementar algumas soluções para a Educação 4.0, que baseia-se no conceito de *Learning by doing*, ou seja aprender fazendo. Esse conceito traz a ideia de que todos nós – principalmente quem está na escola hoje – irá aprender coisas diferentes e de maneiras diferentes, por meio de experiências e projetos. Na Educação 4.0, a análise de desempenho e da aprendizagem podem ser realizadas através de personalização baseada em dados inteligentes. Além do grau de inovação e pesquisa, o projeto visa também a formação de recursos humanos na área de Ciência dos Dados Educacionais.

## 3. Identificação e caracterização do problema

A Educação 4.0 começa a responder às necessidades da “Indústria 4.0” ou da também chamada quarta revolução industrial, onde a linguagem computacional, a Internet das Coisas, a Inteligência Artificial, os robôs e muitas outras tecnologias se somam para dinamizar os processos nos mais diversos segmentos da Indústria. A Educação 4.0 baseia-se no conceito do *Learning by doing*, ou seja aprender fazendo. Utiliza metodologias ativas, para Bacich e Moran (2018), “metodologias ativas são estratégias de ensino centradas na participação efetiva dos estudantes na construção do processo de aprendizagem, de forma flexível e interligada”. O termo 4.0 faz menção ao conceito e uso de Internet inteligente que afirma que os conteúdos destinados aos internautas serão cada vez mais personalizados e interativos. Com esses recursos educacionais, os ambientes computacionais de ensino e os conteúdos digitais passaram a ser vistos como tecnologias essenciais à condução dos processos educacionais; sejam eles presenciais ou à distância.

O uso crescente dessas tecnologias tem gerado grandes volumes de dados, provenientes de interações e de registros contínuos de informações dos professores, alunos, gestores e demais atores dos sistemas educacionais. Extrair novos conhecimento desse volume de dados irá subsidiar na tomada de decisão para os novos desafios da Educação na era digital e esse é o objetivo da Ciência de dados educacionais. A Ciência dos dados educacionais (Educational Data Science em inglês) é

considerada uma das áreas mais promissoras atualmente, e espera-se que nos próximos anos ainda esteja nesse patamar.

Os pilares, que formam a chamada Educação 4.0, estão transformando todas as áreas da Educação. Estamos na era da conectividade global e da facilidade de acesso ao conhecimento. Objetos, dispositivos portáteis, vestíveis e dezenas de outros gadgets estão a cada dia inundando a Internet com novos dados. A evolução exponencial das tecnologias irá permitir que tudo ao nosso redor esteja conectado e que todos esses meios sejam portais de acesso a Educação, gerando uma imensidão de dados.

As técnicas para transformar tais dados em informação é o que coloca a Ciência dos dados educacionais em um patamar da fronteira do conhecimento. O valor imensurável da informação é que irá levar gestores das instituições de ensino públicas e privadas a tomarem suas decisões estratégicas para implantar essa nova era da Educação. Por isso, é fundamental que o país possua uma política estratégica para o desenvolvimento de recursos humanos preparados para o desenvolvimento dessa área de pesquisa.

A Ciência dos dados educacionais é uma área ampla e interdisciplinar, que lida com o desenvolvimento de métodos para explorar dados oriundos de contextos educacionais. Necessita de uma forte base matemática (álgebra linear e estatística), técnicas de inteligência artificial (aprendizado de máquina, stream de dados, tomada de decisão), linguagens de programação (R, Python), banco de dados não relacionais, visualização de dados, frameworks de desenvolvimento (Hadoop, Spark), de tecnologias de infraestrutura de comunicação (nuvem, data center), além das técnicas e metodologias do processo de ensino e da aprendizagem. Assim, um profissional e pesquisador da área de Ciência de dados educacionais precisa possuir um perfil bastante flexível em termos de conhecimento, além de um profundo entendimento sobre o domínio do problema ao qual está lidando.

#### **4. Pressupostos teóricos e justificativa**

Ao longo dos anos, a forma como o processo de ensino e aprendizagem é efetuado nas escolas se modifica. Essas mudanças bruscas aliadas aos avanços tecnológicos são chamadas de Revoluções. A linha do tempo da Educação registra quatro dessas revoluções. Elas estão atreladas as Revoluções Industriais. A Educação 1.0, até meados do século 18, era aquela bem tradicional. O aluno aprende com o professor utilizando livros, cadernos, apostilas, lousa e giz. O ensino é realizado em um local físico definido: a sala de aula. A educação 2.0, partindo da Revolução Industrial até meados do século 20, o aluno, algumas vezes, divide experiências com os colegas, utilizando hardware, software e materiais didáticos, mas o ensino continua sendo realizado em locais definidos, como a sala de aula ou, no máximo, os laboratórios de ciências e de informática. Já na Educação 3.0, a palavra-chave é colaboração. O aluno aprende com o professor e com os seus colegas de forma colaborativa. O ensino começa a ser realizado de forma híbrida, presencial e a distância, com a inserção de múltiplos recursos. E atualmente, na Educação 4.0, o aluno passa a viver a experiência da aprendizagem por meio de projetos colaborativos, nos quais os professores e colegas atuam juntos. Os recursos disponíveis na escola passam a ser usados de maneira criativa e novas estratégias são baseadas nas metodologias ativas para as atividades em sala de aula.

Nessa nova era, em um mundo cada vez mais tecnológico, é necessário que os estudantes tenham um claro entendimento dos princípios e práticas da Ciência da Computação (CC), independente de seus campos de estudo e atuação [CSTA, 2011]. O acesso ao conhecimento dos fundamentos de CC pode auxiliar a utilizar tecnologias de maneira mais consciente, além de apoiá-los na criação de sistemas de computação para melhorar a qualidade de vida das pessoas.

Com todos esses novos cenários produzido pela Educação 4.0 teremos uma mudança no perfil de professores e alunos. Os professores não são mais quem detém exclusivamente o conhecimento, tendo em vista que o seu papel é muito mais amplo e complexo. O professor precisa provocar novos desafios e organizar roteiros personalizados de aprendizagem, tanto individual quanto coletivo. Já os alunos são nativos digitais, imediatista e pouco apegado às fronteiras geográficas. Eles são criadores e adoram testar novas possibilidades de soluções. Eles produzem conhecimento e coloca-os em prática. Para auxiliar de forma eficiente a implantação do conceito da Educação 4.0 se faz necessário a utilização dos conhecimentos da área de Ciência de dados educacionais.

A Ciência de dados educacionais se propõe a explorar esses dados educacionais para o entendimento de situações advindas de ambientes acadêmicos, assim como, propor uma melhor interação entre as áreas de Mineração de dados educacionais, Learning Analytics e Academic Analytics, que são as suas três linhas de pesquisa.

Conceituando cada uma destas linhas, a Mineração de Dados Educacionais (do inglês Educational Data Mining, EDM) é uma área de pesquisa que utiliza as tarefas da Mineração de Dados como Análise Preditiva, Agrupamento e Associação de Dados aplicados a problemas de contexto educacional [Romero e Ventura, 2007; Siemens e Baker, 2012; Costa et al., 2013]. A EDM tem como objetivos fazer descobertas sobre o comportamento dos estudantes e o ambiente no qual a aprendizagem ocorre, fornecendo insumos para o professor ou aluno investigar eventuais padrões descobertos [Romero e Ventura, 2007; Romero et al., 2016; Ducange et al., 2016].

Learning Analytics, por outro lado, foi definido, de acordo com Souza et al (2016), como um “processo para a medição, coleta, análise e comunicação de dados sobre os alunos e os seus contextos, para fins de compreensão e otimização da aprendizagem nos ambientes em que esse processo ocorre” [Siemens et. al., 2011]. Essa área de pesquisa envolve o uso de ferramentas de análise de dados para avaliar processos de aprendizagem estabelecidos por educadores aos seus educandos conforme mostram os trabalhos de Siemens e Baker, (2012), Daniel (2016) e Romero et al. (2016). Nessa mesma linha, qualquer tipo de estratégia de aprendizado, seja com uso de recursos de aprendizagem (objetos de aprendizagem) elaborados pelos educadores ou por uma equipe de TI, ao final espera-se que os resultados sejam analisados e, portanto, é nesse momento em que se insere os estudos de LA [Papamitsiou e Economides, 2014; Martinez-Maldonado et al., 2016; Knight e Littleton, 2016; Quigley et al., 2017].

A Academic Analytics ou AA, tem como foco o uso dos dados oriundos dos sistemas de informação da Instituição de Ensino (IE) para tentar entender os dados cadastrais dos alunos e outros que se relacionam com a vivência acadêmica do aluno na instituição [Campbell e Oblinger, 2007; Baepler e Murdoch, 2010]. Um exemplo é o uso de dados demográficos, desempenho acadêmico, histórico escolar, censo da instituição, uso dos recursos computacionais, financeiro (para IE particulares) e uma série de outros dados que podem implicar de alguma maneira no desempenho do aluno [Campbell e Oblinger, 2007]. No entanto, como colocado por Baepler e Murdoch (2010), a análise

deve ter um olhar mais amplo aos instrutores e alunos, mas também aos gestores. E nesse aspecto, pode-se considerar a AA como apoio aos gestores, a partir do momento em que se usam as análises para avaliar, por exemplo, projetos pedagógicos, processos administrativos, uso dos recursos da biblioteca, entre outros.

Essas áreas de pesquisa começam a se unir em torno de questões e problemas educacionais. Destacando a forma como a comunidade de pesquisadores começou a convergir em torno da mineração de dados educacionais (EDM), Academic Analytics (AA) e, mais recentemente, juntar-se à comunidade de Learning Analytics (LA) para formar um campo atual de pesquisa intitulado Ciência de Dados Educacionais.

Para realizar as análises dos dados, neste projeto, será utilizado a base dados de alunos das turmas iniciais de programação do curso de Ciências e Tecnologia da UFRN, que utilizam um ambiente virtual LOP para execução de exercícios e provas da disciplina de Lógica de Programação.

Essa base de dados foi escolhida, pois, a partir de nossa própria vivência como professores da disciplina de algoritmos e programação, com mais de 10 anos de experiência, observamos os altos índices de reprovações e evasão. Outro motivo relevante para essa escolha é que essa disciplina contém os conceitos básicos da área de Ciência da Computação, que por sua vez, é um dos pilares da implementação da Educação 4.0.

Diante dessa realidade nos preocupamos em saber quais os motivos que levam os alunos a terem tantas dificuldades de aprendizado e consequentemente gerar altos índices de reprovação e evasão dessa disciplina. Como a maioria dos estudos buscam a opinião dos docentes que experienciaram esta disciplina e, a partir das suas avaliações e observações costumam apontar que a desistência/cancelamento da disciplina ocorre pelas seguintes razões: Deficiências relacionadas à expressão em língua materna (escrita), interpretação de textos e enunciados, hábitos de estudo e pesquisa e, especialmente, com formação básica deficitária no que tange a conteúdos de Matemática.

O diferencial deste trabalho será extrair padrões interessantes e implícitos dessa grande quantidade de dados para melhor compreender o comportamento dos estudantes e a forma como eles aprendem. Tais conhecimentos irão servir de subsídio para a melhoria das práticas pedagógicas no ensino em programação, além de ser uma importante ferramenta para viabilizar a personalização do ensino. Buscou-se também identificar quais outros fatores motivacionais, estruturais e socioeconômicos que levaram estes alunos a desistirem do estudo das disciplinas de Programação. Serão analisados dados referentes aos logs de interação de 300 alunos de 3 turmas do curso de Ciência e Tecnologia da UFRN.

## **5. Objetivos e metas**

O projeto tem um caráter de pesquisa e inovação tecnológica. Em aspectos gerais as seguintes metas são desejadas:

1. Formação de recursos humanos na área de Ciências dos Dados Educacionais.
2. Desenvolvimento de soluções para a Educação 4.0 que utilizem inteligência de decisão baseada na Ciência dos Dados.

- a) Desenvolvimento de um modelo com uso das técnicas de Mineração de Dados Educacionais para Recomendações para Estudantes (Apoio ou estudante e feedback).
  - b) Desenvolvimento de um modelo com uso das técnicas de Mineração de Dados Educacionais para Estimativa ou Modelagem de desempenho do aluno, permitindo identificar várias situações, tais como desistências ou reprovações.
  - c) Desenvolvimento de um modelo com uso das técnicas de Mineração de Dados Educacionais para Modelagem de Grupos ou Aprendizado Colaborativo
  - d) Desenvolvimento de um modelo com uso das técnicas de Mineração de Dados Educacionais para Mediação ou Recomendação Pedagógica
  - e) Desenvolvimento de um modelo com uso das técnicas de Mineração de Dados Educacionais para Detecção ou Previsão de Evasão
3. Desenvolvimento de uma solução para visualização de dados que quando empregadas em conjunto com técnicas estatísticas e de aprendizagem de máquina podem potencializar a interpretação dos resultados de dados educacionais.
  4. Disseminação do conhecimento

Em relação aos objetivos específicos, as seguintes ações são almeçadas:

1. Estudo das linguagens de programação R, Python.
2. Estudo sobre a teoria estatística voltada para Ciência dos Dados Educacionais.
3. Estudo sobre técnicas de aprendizado profundo (deep learning em inglês).
4. Estudo sobre técnicas de clusterização e predição.
5. Estudo sobre tecnologias de visualização de dados (bibliotecas e aplicativos).
6. Estudo sobre API (Application Program Interface em inglês) de dados abertos em instituições públicas, incluindo a [api.ufrn.br](http://api.ufrn.br) (API de dados aberto da UFRN).
7. Levantamento de requisitos e projeto da solução de visualização de dados.
8. Proposta novas funcionalidades para o sistema LOP que é uma plataforma de gerenciamento de exercícios para a disciplina de Lógica de Programação da Escola de Ciência e Tecnologia da UFRN.
9. Escrita de artigos em revistas, periódicos e congressos científicos.

## **6. Metodologia e estratégias de ação**

A área da Ciência dos Dados Educacionais exige que os profissionais tenham uma base matemática bastante sólida além de habilidades nas Tecnologias de Informação e Comunicação (TIC) de uma maneira mais ampla. Por isso, é fundamental que seja realizado um estudo bastante profundo sobre as técnicas matemáticas presentes no estado da arte que compõe a grande área de Ciência dos Dados Educacionais. Essa etapa de estudo é essencial para a factibilidade de conclusão das metas estabelecidas no projeto. Além da base de dados científica (Scopus, IEEE, Elsevier, etc), essa etapa inicial irá catalogar e testar as diversas soluções de aplicativos já existentes na literatura e de âmbito livre, que se relacione com a área de Ciência dos Dados Educacionais.

Os estudos iniciais da base matemática serão complementados com os estudos sobre as principais linguagens de programação utilizadas para a área de Ciência dos Dados Educacionais, nomeadamente, R e Python. Cada linguagem tem suas vantagens e

desvantagens. Enumerar e avaliar essas questões é mérito também da respectiva etapa de desenvolvimento.

Como resultado desses dois estudos iniciais, será escrito um artigo para publicação em um periódico descrevendo um Estudo de Revisão (Survey em inglês) sobre Ciências de Dados Educacionais levando consideração aspectos teóricos e práticos.

Paralelo e transversalmente a etapa inicial, estudos sobre a grande área da Educação 4.0 serão realizados. Esses estudos serão bastante importantes para identificar alguns problemas e soluções adotadas no âmbito das TICs, dados abertos e API de desenvolvimento ([api.ufrn.br](http://api.ufrn.br), entre outros). Implementar exemplos práticos que envolvem a implantação dos conceitos da Educação 4.0, com a teoria pesquisada na etapa inicial é de fundamental importância para manter a executabilidade do projeto. Ainda nessa etapa, temas como desempenho dos estudantes, previsão de evasão e recomendações pedagógicas serão estudados com maior profundidade. Um novo artigo de revisão em periódico será utilizado como resultado dessa etapa de execução do projeto. Todavia, diferentemente do primeiro resultado, esse irá explorar aspectos da Ciência dos Dados para a Educação 4.0.

A próxima etapa de execução compõe o levantamento de novos requisitos para a plataforma LOP. Propondo novas ferramentas e soluções para o apoio ao estudante e feedbacks. Essa etapa é bastante densa e envolve também muitas ações de testes de desempenho e execução. As técnicas estudadas na etapa inicial serão utilizadas na estimativa do desempenho do aluno e na previsão de evasão. Dois artigos de periódicos e de conferência serão os resultados dessa etapa.

O aprendizado na elaboração e execução do projeto serão utilizados para a construção de uma solução de visualização de dados. Essa etapa irá finalizar o projeto. Um artigo de periódico e de conferência serão os resultados dessa última meta.

O projeto apresenta um período de 12 meses.

## **7. Recursos existentes para a realização do projeto**

Os Laboratórios Associados Natalnet da UFRN, onde pretende-se fazer o estágio pós-doutoral, possui toda a infraestrutura de redes de computadores com acesso a Internet, para utilizar todos os recursos que a Computação em Nuvem prover para projetos de Ciência de dados.

Os principais recursos serão para coleta de dados, armazenamento de dados, análise de dados e visualização e saída de dados.

- Para a coleta dos dados já temos acesso ao banco de dados da plataforma LOP e os dados incluindo a [api.ufrn.br](http://api.ufrn.br) (API de dados aberto da UFRN).
- Para o armazenamento temos os servidores locais e também fazemos uso um sistema distribuído mais sofisticado (geralmente baseado em nuvem) como o Hadoop.
- Para a análise de dados será utilizada as soluções open source poderosas como as linguagens R e Python.
- Para a visualização e saída de dados iremos utilizar gráficos simples ou ferramentas de visualização como nuvens de palavras que são mais do que suficientes para apresentar informações sobre os dados.

## Cronograma de Atividades

Etapas	Mês de Trabalho											
	01	02	03	04	05	06	07	08	09	10	11	12
Estudo sobre as técnicas matemáticas para Ciência de Dados Educacionais												
Estudos sobre as principais linguagens de programação (R, Python)												
Estudos e Implementação de exemplos práticos da Educação 4.0												
Levantamento de novos requisitos para a plataforma LOP												
Soluções de Visualizações de dados												
Documentação e Divulgação												

## 8. Relevância dos resultados e impactos esperados

A relevância dos resultados do projeto será avaliada em termos dos indicadores de progresso do projeto (ao final do projeto):

- Um modelo computacional para previsão do desempenho dos estudantes, previsão de evasão e recomendações pedagógicas.
- Novos requisitos para a plataforma LOP propondo novas ferramentas e soluções para o apoio ao estudante e feedbacks.
- Publicações de artigos em veículos científicos nacionais e internacionais.
- Relatórios Finais e Artigos submetidos a revistas indexadas.

## 9. Repercussão e/ou impactos dos resultados

Em termos de produto, ao fim do projeto teremos a plataforma LOP com mais funcionalidades para auxiliar no processo de ensino e aprendizagem com o foco nos conceitos de Lógica de Programação. Subsídios e novas estratégias para o estudo e o desenvolvimento de outras metodologias de ensino e técnicas computacionais do mais alto nível para modelar os perfis dos novos alunos e personalizar o ensino para cada perfil. Espera-se que essa plataforma LOP tenha um grande impacto no grupo Natalnet (UFRN), pois poderão ser utilizadas como ambientes de testes nas futuras pesquisas na área de Ciência de dados Educacionais e do ensino e aprendizado dos conceitos de Lógica de Computação, por professores e alunos que necessitem deste tipo de suporte experimental prático.



Como pretendemos disponibilizar todo o conhecimento gerado para a comunidade de Ciência de dados Educacional do Brasil, espera-se que a partir desta plataforma, novos projetos na área possam ser definidos, submetidos e implementados não só em convocatórias nacionais como também internacionais. Com este projeto, o grupo de ciência de dados Educacionais da UFRN será capacitado e ainda terá a capacidade de desenvolver novos projetos que complementam os desenvolvidos até o momento.

## **10.Referências bibliográficas**

BACICH, L; MORAN, J. Metodologias Ativas para uma Educação Inovadora: Uma Abordagem Teórico Prática. Penso Editora, 2017.

Baepler, P., & Murdoch, C. J. (2010). Academic analytics and data mining in higher education. *International Journal for the Scholarship of Teaching and Learning*, 4(2), 17.

Campbell, J. P., & Oblinger, D. G. (2007). Academic analytics. *EDUCAUSE review*, 42(4), 40-57.

Costa, E., Baker, R. S., Amorim, L., Magalhães, J., & Marinho, T. (2013). Mineração de Dados Educacionais: Conceitos, Técnicas, Ferramentas e Aplicações. *Jornada de Atualização em Informática na Educação*, 1(1), 1-29.

CSTA - Computer Science Teacher Association (2011). CSTA K-12 Computer Science Standards. CSTA Standards Task Force. ACM - Association for Computing Machinery.

Daniel, B. K. (Ed.). (2016). *Big Data and Learning Analytics in Higher Education: Current Theory and Practice*. Springer.

Ducange, P., Pecori, R., Sarti, L., & Vecchio, M. (2016, October). Educational Big Data Mining: How to Enhance Virtual Learning Environments. In *International Conference on European Transnational Education* (pp. 681-690). Springer International Publishing.

Knight, S., & Littleton, K. (2016). Dialogue as Data in Learning Analytics for Productive Educational Dialogue. *Journal of Learning Analytics*, 2(3), 111-143.

Martinez-Maldonado, R., Schneider, B., Charleer, S., Shum, S. B., Klerkx, J., & Duval, E. (2016, April). Interactive surfaces and learning analytics: data, orchestration aspects, pedagogical uses and challenges. In *Proceedings of the Sixth International Conference on Learning Analytics & Knowledge* (pp. 124-133). ACM.

Papamitsiou, Z. K., & Economides, A. A. (2014). Learning analytics and educational data mining in practice: A systematic literature review of empirical evidence. *Educational Technology & Society*, 17(4), 49-64.

Quigley, D., Ostwald, J., & Sumner, T. (2017, March). Scientific modeling: using learning analytics to examine student practices and classroom variation. In *Proceedings of the Seventh International Learning Analytics & Knowledge Conference* (pp. 329-338). ACM.

Romero, C., Cerezo, R., Bogarín, A., & Sánchez-Santillán, M. (2016). EDUCATIONAL PROCESS MINING. *Data Mining and Learning Analytics: Applications in Educational Research*, 1-28.

Romero, C.; Ventura, S. (2007). Educational data mining: A survey from 1995 to 2005. *Expert systems with applications*, Elsevier, 33(1), 135–146.

Siemens, G.; Baker, R. S. (2012). Learning analytics and educational data mining: towards communication and collaboration. In: ACM. Proceedings of the 2nd international conference on learning analytics and knowledge (p. 252–254).

Siemens G. LAK'11 1st International Conference on Learning Analytics and Knowledge. Disponível em < <https://tekri.athabascau.ca/analytics/>> Acesso em: 10 de janeiro de 2019.