



XXIX CONGRESSO DE INICIAÇÃO CIENTÍFICA (CIC)  
2019

UACSA, UAST, UFAPE, CODAI e UEADTEC  
Universidade Federal Rural de Pernambuco  
Pró-Reitoria de Pesquisa e Pós-Graduação  
Coordenação de Programas Especiais



## PROCESSAMENTO DISTRIBUÍDO PARA DESENVOLVIMENTO DE MODELOS DESCRITIVOS E PREDITIVOS DO APRENDIZADO DE MÁQUINA

Francisco Queiroga da Silva Neto<sup>1</sup>, Romero Claudino Ferreira da Silva<sup>1</sup>, Roberta Macêdo Marques Gouveia<sup>2</sup>  
E-mail: franciscofqueiroga@gmail.com

1 UFRPE/Discente

2 UFRPE/Docente

As atividades desenvolvidas, assim como os resultados obtidos para o plano de trabalho “Processamento Distribuído para Desenvolvimento de Modelos Descritivos e Preditivos do Aprendizado de Máquina” fazem parte do projeto “Computação Distribuída e Aprendizado de Máquina Aplicados em Grandes Volumes de Dados Educacionais”. O plano de trabalho faz parte do Programa Institucional de Bolsas de Iniciação Científica (PIBIC), que propõe a aplicação de técnicas, assim como a pesquisa instrumental baseada em modelos de sistemas distribuídos que auxiliem na manipulação de grandes volumes de dados. O Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (INEP) foi a fonte dos dados utilizados no projeto em questão. Anualmente, o INEP é responsável pela realização e divulgação de sinopses estatísticas e microdados que descrevem diversas informações sobre a educação brasileira, de nível básico e superior. Utilizar dados deste tipo na aplicação e avaliação de modelos de predição de Aprendizado de Máquina enriquece o processo de obtenção de informações relevantes, foco principal da Mineração de Dados Educacionais. Entretanto, são escassas as ferramentas, assim como técnicas para o manuseio de grandes volumes de dados educacionais. A dificuldade na obtenção de metodologias que enfatizem ou apliquem medidas que viabilizem o manuseio de grandes de dados educacionais é um trabalho relevante e de grande importância à ser expandido. O desenvolvimento do projeto identifica e aplica técnicas e ferramenta baseados em Computação em Nuvem para construção de modelos de extração e descoberta de conhecimento, como o Knowledge-Discovery in Databases (KDD). O desenvolvimento do projeto deu-se através de pesquisas em soluções voltadas para Nuvem, manuseio de dados com ênfase na fase de Mineração de Dados (MD) e análises voltadas para a definição de possíveis cenários que mensuram o desempenho de cada modelo de Machine Learning aplicado e sua relevância para a melhoria no processo de descoberta de conhecimento.

**Palavras-chave:** cloud computing, machine learning, data mining, knowledge-discovery in databases, big data, data science.

**Área do Conhecimento:** Ciências Exatas e da Terra.

Realização:



Apoio:



FUNDAÇÃO APOLÔNIO SALLES  
F A D U R P E