

# V WORKSHOP EM CIÊNCIA DE DADOS

JORNADA DE CIÊNCIA DE DADOS – PROGRAMA DE VERÃO DO LNCC 2018

Auditório A, LNCC/MCTIC, Petrópolis, RJ, 09 de março de 2018

Organização: Fabio Porto (LNCC), Eduardo Ogasawara (CEFET-RJ), Artur Ziviani (LNCC)

09:00-09:30	<b>Abertura e Visão Geral das Atividades de Ciência de Dados no LNCC</b> Artur Ziviani (LNCC) e Fabio Porto (LNCC)
09:30-10:30 <b>Keynote 1</b>	<b>Ciência de Dados e Aplicações em Bancos</b> Renner Pereira de Castro (Itau-Unibanco)
10:30-11:00	<b>COFFEE BREAK</b>
11:00-12:30 <b>Palestras Convidadas</b>	<b>Arcabouço computacional escalável para modelagem de nicho ecológico: Aplicação à avaliação de distribuição de espécies para a flora Brasil 2020</b> Marinez Ferreira de Siqueira (Jardim Botânico do Rio de Janeiro) <b>Artificial Neural Networks: Recent Advances, Current Trends and Open Problems</b> Eduardo Bezerra (CEFET-RJ) <b>Deep Learning for Structured Data, with Applications on Oil and Gas</b> Aline Paes (UFF)
12:30-14:00	<b>ALMOÇO</b>
14:00-15:00 <b>Keynote 2</b>	<b>Desafios e oportunidades de pesquisa relacionados à tecnologia blockchain</b> Daniel Sadoc Menasché (UFRJ)
15:00-16:00 <b>Painel</b>	<b>Painel sobre Experiências, Desafios e Perspectivas de Inteligência Artificial</b> Moderador: Eduardo Ogasawara (CEFET-RJ)
16:00-16:15	<b>Encerramento</b>

Promoção:



<http://cid.lncc.br>

Agradecimentos:



**Laboratório Nacional de Computação Científica (LNCC)**

Av. Getúlio Vargas, 333 – Quitandinha – 25651-075 – Petrópolis, RJ

## **Keynote 1:**

**Renner Pereira de Castro**

Título:

**Ciência de Dados e Aplicações em Bancos**

Resumo:

Esta palestra abordará uma visão geral de Ciência de dados no Itaú Unibanco, abordando atividades funcionais do dia a dia de um cientista de dados e desafios de aprendizado de máquinas nas áreas do banco. Será descrito em detalhes alguns problemas reais de negócio, descrevendo os desafios desde a aquisição de dados, treinamento de modelos e implantação em produção.

Biografia:

Renner Castro é Principal Data Scientist no Itaú Unibanco. Adquiriu em 2008 o título de Doutor em Matemática Aplicada com ênfase em Computação Gráfica pela PUC-Rio, onde foi professor adjunto nos anos de 2006 e 2007. Adquiriu em 2004 o título de Mestre em Matemática Pura pelo Instituto Nacional de Matemática Pura e Aplicada – IMPA. É formado em Bacharel em Matemática pela Universidade Federal de Uberlândia – UFU. Suas principais atividades atuais são de liderar projetos no Itaú Unibanco relacionados a Inteligência Artificial, treinamento e seleção de cientistas de dados.

---

## **Keynote 2:**

**Daniel Sadoc Menasché (UFRJ)**

Título:

**Desafios e oportunidades de pesquisa relacionados à tecnologia blockchain**

Resumo:

A tecnologia blockchain está revolucionando diversas áreas de negócio e do conhecimento. Como toda tecnologia disruptiva, ela traz os seus desafios e oportunidades de pesquisa. Nesta palestra, iremos abordar alguns desafios relacionados à tecnologia blockchain, no que tange 1) eficiência e escalabilidade, 2) privacidade, e 3) detecção de anomalias. Iremos apontar conexões entre tais desafios e ferramentas de modelagem e avaliação de desempenho, e aprendizado por máquina. Desta forma, visamos promissoras oportunidades de pesquisa na interseção entre ciência da computação e blockchains.

Biografia:

Daniel S. Menasché formou-se em Ciência da Computação pela Universidade Federal do Rio de Janeiro (UFRJ), e obteve os títulos de Mestrado pelo Programa de Engenharia de Sistemas e Computação da COPPE/UFRJ, e Doutorado pela Universidade de Massachusetts em Amherst (EUA), em 2005 e 2011, respectivamente. Desde 2011, é professor adjunto do Departamento de Ciência da Computação do Instituto de Matemática da UFRJ. Foi eleito membro afiliado da Academia Brasileira de Ciências, no período de 2016 a 2020. Seus interesses de pesquisa envolvem modelagem e análise de sistemas de computação e comunicação, segurança e aprendizado por máquina.

## **Palestrante 1:**

**Marinez Ferreira de Siqueira (Jardim Botânico do Rio de Janeiro)**

Título:

**Arcabouço computacional escalável para modelagem de nicho ecológico: aplicação à avaliação de distribuição de espécies para a flora Brasil 2020**

Resumo:

Ferramentas de análise e síntese de dados de biodiversidade, a exemplo da modelagem de nicho ecológico, normalmente empregam diversas aplicações e bibliotecas computacionais distintas que precisam ser compostas e encadeadas durante a sua execução, sendo um caso típico para a utilização de sistemas de gerenciamento de workflows científicos. Tais processos são computacionalmente demandantes, o que torna importante o uso de ferramentas que sejam escaláveis. Um recurso disponível para a resolução deste problema é o ambiente de análise de dados R, que apresenta várias possibilidades de automatização de processos, mas requer conhecimentos avançados de programação e otimização para a obtenção dos resultados desejados neste processo. Um dos principais objetivos deste trabalho é criar um arcabouço de modelagem, que disponibilize os processos e funções R, que permitam a execução das diversas etapas descritas no processo de aquisição, limpeza de dados, seleção de variáveis preditoras, parametrização de algoritmos e pós-processamento de modelos, superando as barreiras mais comuns de processamento e performance, fornecendo novas abordagens, para um público alvo definido, quanto a disponibilização de resultados em vários formatos, visando a integração com várias etapas de pós-análise, tais como diferentes processos envolvendo impactos de mudanças climáticas e antrópicas na distribuição de espécies.

Biografia:

Marinez Ferreira de Siqueira possui graduação em Ciências Biológicas pela Universidade Estadual de Campinas (1988), mestrado em Ecologia pela Universidade Estadual de Campinas (1994) e doutorado em Ciências da Engenharia Ambiental pela Universidade de São Paulo (2005). Atualmente é Pesquisador Associado do Instituto de Pesquisas do Jardim Botânico do Rio de Janeiro. Tem experiência nas áreas de Ecologia Vegetal, Modelagem Ecológica, Sistemas de Informação e Banco de Dados, atuando principalmente em modelagem de nicho ecológico e distribuição de espécies em diferentes formações vegetais. Tem experiência no uso de Sistemas de Informação Geográfica para análises espaciais envolvendo distribuição de espécies.

---

## **Palestrante 2:**

**Eduardo Bezerra (CEFET-RJ)**

Título:

**Artificial Neural Networks: Recent Advances, Current Trends and Open Problems**

Resumo:

Research in Artificial Neural Networks (ANN) aim at devising techniques to simulate the behavior of the human brain in tasks such as visual recognition, speech recognition and natural language processing. Despite being an old area of research, the field has witnessed a surge of interest (both from academia and general public) in the past years, due to recent breakthroughs. In this lecture, we will review some recent advancements which led to the so called ANN renaissance. We will also present some current research trends and open problems.

Biografia:

Eduardo Bezerra has been a professor at the Informatics Department of the Federal Center for Technological Education of Rio de Janeiro (CEFET/RJ) since 2005. He studied Computer Science in UFRJ from 1992 to 1994 and obtained his Ph.D. degree in computer science at COPPE/UFRJ in 2006. He is a permanent lecturer and founding member at the Computer Science Graduate Program of CEFET/RJ since 2016. He is the author of two books and several scientific research papers. His research interests are intelligent data analysis, machine learning, and AI.1

**Palestrante 3:**  
**Aline Paes (UFF)**

Título:

**Deep Learning for Structured Data, with Applications on Oil and Gas**

Resumo:

The AI sub-area of Learning Representations aims at developing techniques that automatically find useful representations of data so that Machine Learning becomes less dependent on the choice and design of specific features. Learning meaningful and useful representations from data, usually following an end-to-end approach, has been successfully progressed by employing Deep Learning methods. In the last five years, Deep Learning techniques have rapidly achieved impressive results in the state-of-the-art in non-structured domains, such as image, natural language, sound, and speech. However, a large amount of information is also represented and stored as structured data, in such a way that the relationships and hierarchies surrounding the objects within them should not be ignored. In this talk, we will present recent results on learning representations for structured data such that it becomes also possible to learn from them using the full potential of Deep Learning methods. We will discuss the applications of such methods in problems of the industry of Oil and Gas.

Biografia:

Aline Paes is an assistant professor in the Institute of Computing and PPGC, Universidade Federal Fluminense. She holds a D.Sc. and an M.Sc. in Systems Engineering and Computer Science, both received from PESC/COPPE-UF RJ. As a Ph.D. candidate, she was a visiting student researcher at the Department of Computing, Imperial College London for one year. Her research interests are mainly in Statistical and Relational AI (Domain Adaptation and Transfer Learning, Online and Reinforcement Learning) and its connections with Learning of Representations with Deep Methods. She tries to employ the techniques developed in the aforementioned areas in applications of Web Knowledge Data, Social Media and Urban Data, Scientific Workflows, Storytelling, Natural Language Processing and Computer Vision. She is interested in Artificial Intelligence in general, and also in employing AI techniques in other areas of Computer Science.