



DAL & COMUNIDADE: UM RELATO DE EXPERIÊNCIA EM DIFUSÃO DE CIÊNCIA DE DADOS

Eduardo Ogasawara
eduardo.ogasawara@cefet-rj.br
<https://eic.cefet-rj.br/~eogasawara>

Biografia

- Doutor em Engenharia de Sistemas e Computação (COPPE/UFRJ)
- Docente no CEFET/RJ
 - Bacharelado em Ciência da Computação
 - Programa de Pós-Graduação em Ciência da Computação (PPCIC)
 - Programa de Pós-Graduação em Eng. de Produção e Sistemas (PPPRO)
- Membro IEEE (Sênior), SBC e ACM
- Líder do Data Analytics Lab

*Grupo de Pesquisa do CEFET/RJ
formalizado no CNPq*



**Transformando dados em conhecimento.
Ciência de Dados em ação.**

Contexto da criação do DAL

- Origem do Grupo de Pesquisa em Computação Aplicada (GPCA) criado em 2012
- Em 2016 foi feito o spin-off do GPCA e passou a se chamar Grupo de Pesquisa em Ciência de Dados (GPCD)
- Em 2019 passou a se chamar de Data Analytics Lab (DAL)
- Missão do DAL:
 - Desenvolver métodos e ferramentas de análise de dados, com foco em séries temporais e ambientes não estacionários, promovendo pesquisa aplicada, software de acesso aberto e difusão de conhecimento para a comunidade acadêmica e a sociedade
- Visão:
 - Tornar-se referência nacional e internacional em ciência de dados para séries temporais, integrando produção científica, inovação de software e impacto social por meio da difusão e aplicação prática em diversos domínios

Principais temas abordados pelo Data Analytics Lab

- Mineração de Dados
 - Pré-processamento de dados
 - Padrões frequentes
 - Predição (classificação e regressão)
- Análise de séries temporais
 - Pré-processamento de dados
 - Predição
 - Detecção de eventos
 - Detecção de anomalias
 - Detecção de pontos de mudança e de desvios de conceito
 - Descoberta de motivos
 - Detecção online
 - Avaliação de detecções



Equipe

Pesquisadores



Eduardo
Ogasawara



Fabio Porto



Helga Balbi



Kele Belloze



Rafaelli Coutinho

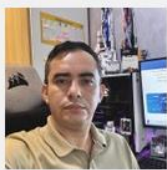


Rebecca Salles

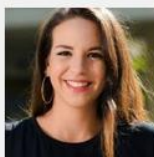
Doutorandos



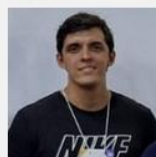
Ellen Paixão



Janio Lima



Lais Baroni



Lucas Giusti*



Marlon Mesquita

Mestrandos



Aline da Silva



Edson Sobrinho



Eduardo dos
Santos



Igor Andrade



Josélia Rabelo



Luiz Oliveira



Michel Reis*



Priscilla da Costa



Ricardo Castro



Rodrigo Machado

Egressos e status atual

- Concluídas
 - 5 Doutorados
 - 25 Mestrados
 - 32 Trabalhos de conclusão de curso (TCC)
 - 34 Iniciações científicas (IC)
- Em andamento
 - 6 Doutorandos
 - 11 Mestrandos
 - 5 TCCs
 - 1 IC

Produção científica do grupo

- 185 artigos publicados, sendo 55 em periódicos científicos

Métricas do Scopus

1,449

Citations by 1,030 documents

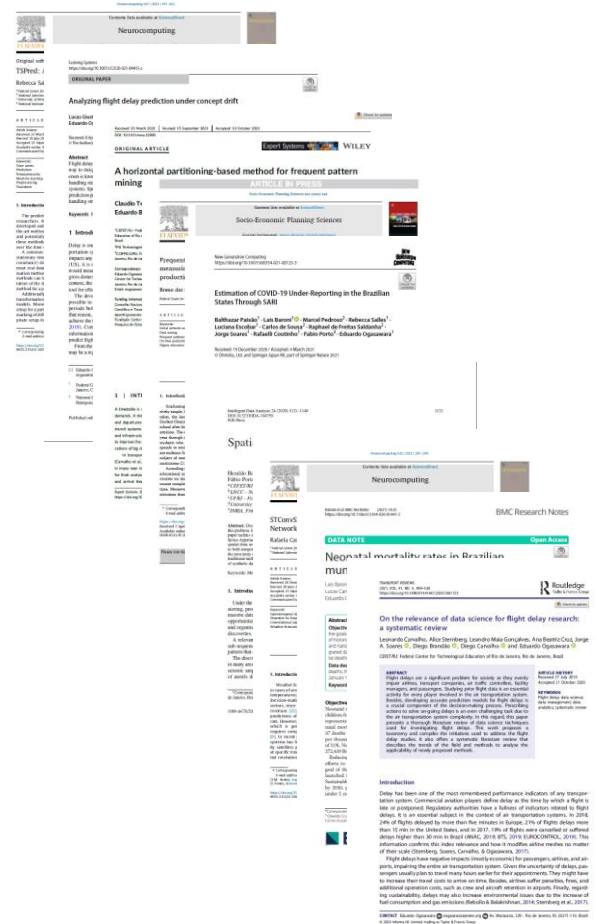
106

Documents

19

h-index

Document & citation trends





DAL Toolbox

Stars 4 downloads 1102/month

The goal of DAL Toolbox is to provide a series data analytics functions organized as a framework. It supports data preprocessing, classification, regression, clustering, and time series prediction functions.

Examples

Graphics: <https://github.com/cefet-rj-dal/daltoolbox/tree/main/graphics/>

Transformation: <https://github.com/cefet-rj-dal/daltoolbox/tree/main/transf/>

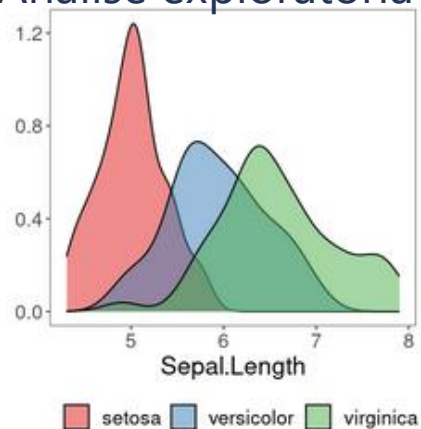
Classification: <https://github.com/cefet-rj-dal/daltoolbox/tree/main/classification/>

Clustering: <https://github.com/cefet-rj-dal/daltoolbox/tree/main/clustering/>

Regression: <https://github.com/cefet-rj-dal/daltoolbox/tree/main/regression/>

Time series: <https://github.com/cefet-rj-dal/daltoolbox/tree/main/timeseries/>

Análise exploratória



Exemplo de classificação

```
# Model testing
test_prediction <- predict(model, iris_test)

iris_test_predictand <- adjust_class_label(iris_test[, "Species"])

# Evaluating # setosa as primary class
test_eval <- evaluate(model, iris_test_predictand, test_prediction)
print(test_eval$metrics)
```

```
## accuracy TP TN FP FN precision recall sensitivity specificity f1
## 1 0.9666667 11 19 0 0 1 1 1 1
```

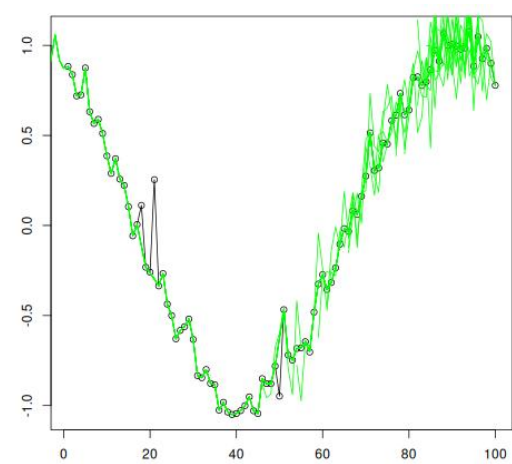
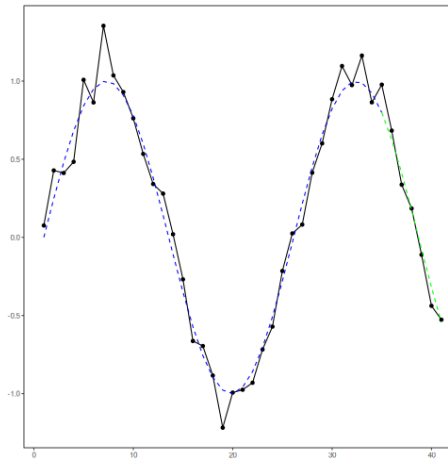
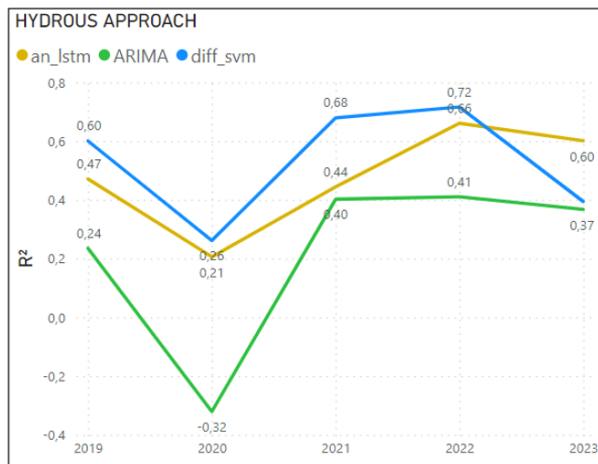


Stars 5 downloads 19K/month

TSPredIT (Time Series Prediction with Integrated Tuning) is a framework for time series prediction with automatic preprocessing and hyperparameter optimization. It is built on top of the [DAL Toolbox](#) and enhances its capabilities by integrating several advanced functionalities:

- Automatic hyperparameter tuning for models and preprocessing
- Outlier detection and removal
- Time series data augmentation
- Filtering techniques for noise reduction
- Ensemble learning support
- Modular and extensible workflow for predictive modeling

TSPredIT is designed to provide a **more flexible and customizable pipeline** for building predictive models on time series data, making it easier to compare alternatives and automate repetitive tasks.





Stars 22 downloads 18K/month

Harbinger is a framework for event detection in time series. It provides an integrated environment for anomaly detection, change point detection, and motif discovery. Harbinger offers a broad range of methods and functions for plotting and evaluating detected events.

For anomaly detection, methods are based on: - Machine learning model deviation: Conv1D, ELM, MLP, LSTM, Random Regression Forest, and SVM - Classification models: Decision Tree, KNN, MLP, Naive Bayes, Random Forest, and SVM - Clustering: k-means and DTW - Statistical techniques: ARIMA, FBIAD, GARCH

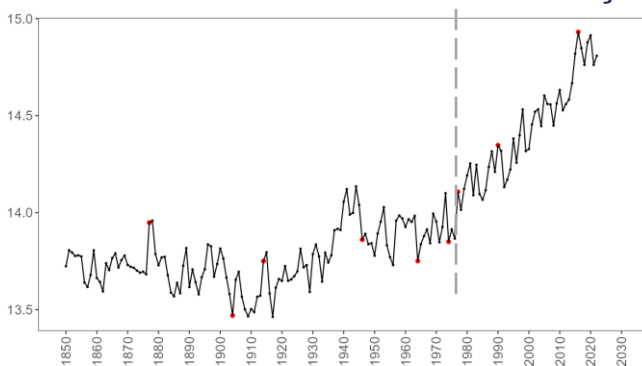
For change point detection, Harbinger includes: - Linear regression, ARIMA, ETS, and GARCH-based approaches - Classic methods such as AMOC, ChowTest, Binary Segmentation (BinSeg), GFT, and PELT

For motif discovery, it provides: - Methods based on Hashing and Matrix Profile

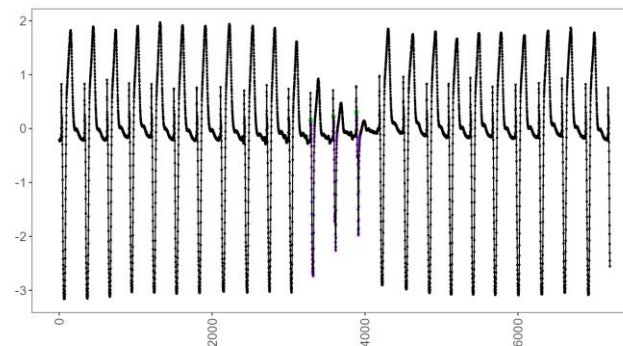
Harbinger also supports **multivariate time series analysis** and **event evaluation** using both traditional and soft computing metrics.

The architecture of Harbinger is based on **Experiment Lines** and is built on top of the [DAL Toolbox](#). This design makes it easy to extend and integrate new methods into the framework.

Anomalia e Ponto de Mudança



Descoberta de Motifs



Outros artefatos computacionais



Hanafuda

GPCA Cartas

Este app é compatível com todos os seus dispositivos.



ipeadata

BRASIL Acesso à Informação Participe Serviços Legislação Canais

macroeconômico regional social

macroeconômico regional social

Dados econômicos e Dados econômicos Dados e indicadores

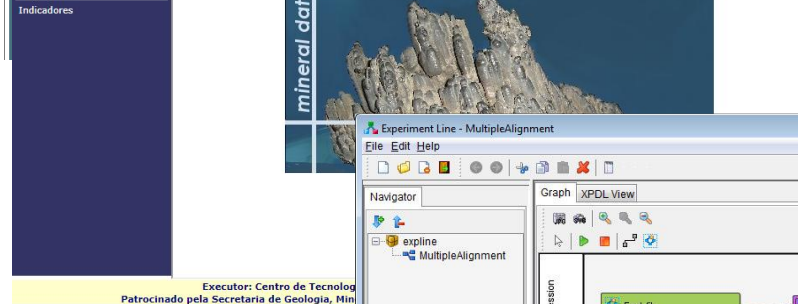


mineral data

Séries Históricas do Setor Mineral Brasileiro

Home Mapa do Site Fale conosco

Mineral Data Substâncias Temáticas Fontes Catálogo Indicadores



Experiment Line - MultipleAlignment

Fast file MultipleAlign External Execution Display

TSPred: Functions for Benchmarking Time Series Prediction

Functions for defining and conducting a time series prediction process including pre(post)processing, decomposition, modelling, prediction and accuracy assessment. The generated models and its yielded prediction errors can be used for benchmarking other time series prediction methods and for creating a demand for the refinement of such methods. For this purpose, benchmark data from prediction competitions may be used.

Version: 5.1.1

gstsm: Generalized Spatial-Time Sequence Miner

Instalado

Implementations of the algorithms present article Generalized Spatial-Time Sequence Miner, original title (Castro, Antonio; Borges, Heraldo; Pacitti, Esther; Porto, Fabio; Coutinho, Rafaeli; Ogasawara, Eduardo. Generalização de Mineração de Sequências Restritas no Espaço e no

STMotif: Discovery of Motifs in Spatial-Time Series

Allow to identify motifs in spatial-time series. A motif is a previously unknown subsequence of a (spatial) time series with relevant number of occurrences. For this purpose, the Combined Series Approach (CSA) is used.

Version: 2.0.2

Imports: stats, ggplot2, reshape2, scales, grDevices, RColorBrewer

Suggests: knitr, rmarkdown, testthat

Published: 2024-02-23

DOI: 10.32614/CRAN.package.STMotif

Author: Heraldo Borges [aut, cre] (CEFET/RJ), Amin Bazaz [aut] (Polytech'Montpellier), Esther Pacitti [aut] (INRIA/Polytech'Montpellier), Eduardo Ogasawara [aut] (CEFET/RJ)

Maintainer: Heraldo Borges <stmotif@eic.cefet-rj.br>

BugReports: <https://github.com/heraldoborges/STMotif/issues>

License: MIT + file LICENSE

URL: <https://github.com/heraldoborges/STMotif/wiki>

NeedsCompilation: no

Materials: NEWS

CRAN checks: [STMotif results](#)

lirmm-00806557, version 1

Chiron: A Parallel Engine for Algebraic Scientific Workflows

Eduardo Ogasawara¹, Dias Jonas¹, Vitor Silva¹, Chirigati Fernando¹, Oliveira Daniel De¹, Fabio Porto², Marta Mattoso¹, Patrick Valduriez^{3,4} [Details](#)

- 1 COPPE-UF RJ - COPPE - Instituto Alberto Luiz Coimbra de Pós-Graduação e Pesquisa de Engenharia
- 2 LNCC / MCT - Laboratório Nacional de Computação Científica (Rio de Janeiro)
- 3 ZENITH - Scientific Data Management
- 4 LIRMM - Laboratoire d'Informatique de Robotique et de Microélectronique de Montpellier, CRISAM - Inria Sophia Antipolis - Méditerranée

Abstract: Large-scale scientific experiments based on computer simulations are typically modeled as scientific workflows, which eases the chaining of different programs. These scientific workflows are defined, executed and monitored by Scientific Workflow Management Systems (SWIMS). As these experiments manage large amounts

IDENTIFIERS

- HAL Id: lirmm-00806557, version 1

COLLECTIONS

INRIA | IBC | INRA | ZENITH | LIRMM | MIPS | UNIV-MONTPELLIER

CITATION

Eduardo Ogasawara, Dias Jonas, Vitor Silva, Chirigati Fernando, Oliveira Daniel De, et al. Chiron: A Parallel Engine for Algebraic Scientific Workflows. *Concurrency and Computation: Practice and Experience*, Wiley, 2013, 25 (16), pp.2327-2341. (lirmm-00806557)

Desafios de difusão de Ciência de Dados

- Dificuldade de acesso a ferramentas
- Pouca familiaridade de discentes/egressos com métodos
- Escassez de materiais em português
- Necessidade de dialogar com setores externos (governo, sociedade, empresas)

Estratégias usadas pelo DAL

- Foco nos temas: Séries Temporais
- Ferramentas abertas: DAL Toolbox, TSPredIT, Harbinger
- Difusão de Conhecimento (Youtube):
 - Disciplinas
 - Tutoriais de R e Linux
 - DAL Talks
 - DAL Lives
- Jornada de Ciência de Dados (em parceria com LNCC)
- Apresentação de resultados de colaborações

Difusão de conhecimento

- Metodologia Científica
 - <https://eic.cefet-rj.br/~eogasawara/metodologia-cientifica/>
- Mineração de Dados
 - <https://eic.cefet-rj.br/~eogasawara/data-mining/>
- Banco de Dados
 - <https://eic.cefet-rj.br/~eogasawara/banco-de-dados/>
- Programação em R
 - <https://eic.cefet-rj.br/~eogasawara/tutorial-r/>
- Outros assuntos
 - Linux Básico para Data Analytics
 - Apresentações



Canal do DAL no Youtube ***(<https://www.youtube.com/@dataanalyticslab>)***

- DAL Lives (2024-2025)



Atrasos de Voos no Brasil: Análise Pioneira com Padrões Frequentes ⋮

138 visualizações • Transmitido há 4 meses

- DAL Talks (2021-2025)



Planejamento e Execução da Pesquisa ⋮

165 visualizações • há 2 anos

- Computação em Foco (2023)



Participações na SEPEX

- 2022: Ciência de Dados nos Eixos Ciências, Empresas e Governos
- 2023: Gerência de Dados, Modelos e Configuração no Contexto de Inteligência Artificial
 - Sistema de Consulta a Edições Especiais
- 2024: Painel de Ferramentas Baseadas em Inteligência Artificial para Apoio à Metodologia de Pesquisa
- 2025: DAL & Comunidade: Um Relato de Experiência em Difusão de Ciência de Dados

Sistema de Consulta a Edições Especiais

The screenshot displays a web browser window with the URL `rukbat.eic.cefet-rj.br/special/`. The page has a dark blue header with the text "Special Issues" and "Ciência da computação". A search bar with the placeholder "Pesquisar palavras-chave" is located in the header. On the left side, there is a vertical sidebar with icons for home, search, and other functions. The main content area features three white boxes, each representing a special issue. Each box contains a title, a submission deadline, and a brief description or list of guest editors.

Special Issues
Ciência da computação

Pesquisar palavras-chave

Selected Papers from the 20th Wireless On-demand Network systems and Services Conference
Prazo para submissão: 01 April 2025
WONS is a small, single track, selective conference that follows and often anticipate the novelties in "On-Demand Networking", i.e., all the topics related to wireless networking outside the main stream of cellular networks and services that are provided where and when they are needed. The Guest Edi...

Reliable and Secure Large Language Models for Software Engineering
Prazo para submissão: 01 April 2025
Dr. Gabriele Bavota, Università della Svizzera italiana, Switzerland Dr. Pietro Liguori, University of Naples Federico II, Italy, Email: pietro.liguori@unina.it Dr. Antonio Mastropaolo, William & Mary, USA Special issue information:...

Security and Privacy in AIoT-enabled Smart Society
Prazo para submissão: 01 April 2025
Associate Prof. Qin Liu Hunan University, Changsha, China Prof. Kouichi Sakurai Kyushu University, Fukuoka, Japan Prof. Richard Hill University of Huddersfield, Huddersfield, UK Associate Prof. Wenjia Li UK New York Institute of Technology, Old Westbury, USA Special issue information:...

Próximos passos

- Expandir difusão para escolas públicas de ensino médio
- Consolidar atuação com o setor público (Prefeitura do Rio, Fiocruz)
- Preparar materiais acessíveis para as ferramentas



DAL & COMUNIDADE: UM RELATO DE EXPERIÊNCIA EM DIFUSÃO DE CIÊNCIA DE DADOS

Eduardo Ogasawara
eduardo.ogasawara@cefet-rj.br
<https://eic.cefet-rj.br/~eogasawara>