



# GT7: Técnicas Avançadas de Análise

GUSTAVO GIL, THIAGO TOMEI

UFRGS, SPRACE-Unesp

# Equipe



Coordenação: Gustavo Gil (UFRGS) e Thiago Tomei (SPRACE-Unesp)



Claudio Lenz (UFRJ)

Antonio Vilela (UERJ)  
Cesar Bernardes (UFRGS)  
Patricia Telles (CBPF)



Cristiane Jahnke (USP)

ALICE

Erica Policarpo (UFRJ)  
Murilo Rangel (UFRJ)



Marcelo Fernandes (UFRN)  
Marisilvia Donadelli (USP)  
Yara Coutinho (UFRJ)

<https://indico.cern.ch/category/17270/>

# Objetivo

*Disseminar o conhecimento sobre ferramentas de análise (e.g. técnicas de aprendizado de máquina), manipulação de dados com estruturas heterogêneas (e.g. RDataFrame) e interoperabilidade com o ecossistema de ciência de dados. Fomentar a troca de experiências no desenvolvimento de frameworks para análises em nuvem (CERN SWAN, Google Colab, etc) com foco em portabilidade e reprodutibilidade, e **ênfase nos membros mais jovens.***

# Execução Orçamentária (I Workshop)

Recursos previstos no projeto:

**R\$ 217.000,00**

- Ano 3: Primeira edição da Escola de treinamento presencial
- Ano 4 ou 5\*: Segunda edição da Escola de treinamento presencial
  - Em se obtendo sucesso em organizar uma edição da Escola com recursos externos, o dispêndio pode ser prorrogado para parcela posterior dos recursos do projeto no Ano 5.



## Budget

Workshops	R\$ 266.000,00
Partic. Conferências	R\$ 396.100,00
Missões CERN	R\$ 977.000,00
Bolsas	R\$ 540.000,00
GT - Si	R\$ 1.052.100,00
GT - Gas	R\$ 1.027.000,00
GT - CF	R\$ 280.000,00
GT - Eletrônica	R\$ 361.597,80
GT - Téc. Associadas	R\$ 118.400,00
GT - Infra Comput.	R\$ 873.273,69
GT - Téc. Análises	R\$ 217.000,00
Divulgação Cientif.	R\$ 306.840,00

1/3

2/3

*Palestra Prof. Bediaga, 01/jul/2024*



# Escola INCT de Análise de Dados 2024

4–8 de nov. de 2024  
SPRACE, NCC-UNESP  
Fuso horário America/Sao\_Paulo

Digite o seu termo de pesquisa:

## Visão Geral

### Ferramentas

Local

Tabela de Horários

Lista de Contribuição

Registro

Lista de participantes

### Contact

[escola-inct.brasil@cern.ch](mailto:escola-inct.brasil@cern.ch)



## Escola INCT de Análise de Dados 2024



SPRACE



Núcleo de  
Computação  
Científica



Como parte das atividades do INCT CERN Brasil, a Escola INCT de Análise de Dados 2024, organizada pelo Grupo de Trabalho de Técnica Avançadas de Análise de Dados (GT7), ocorrerá entre os dias 4 e 8 de novembro de 2024 com o objetivo de proporcionar a estudantes de graduação e pós-graduação a nível de Mestrado os conhecimentos básicos necessários para analisar dados provenientes de detectores de partículas. Esta Escola, hospedada pelo São Paulo Regional Analysis Center (SPRACE), ocorrerá nas dependências do Núcleo de Computação Científica (NCC) da UNESP, o qual conta com a infraestrutura computacional de um Tier-2 do CERN voltado à Colaboração CMS. O programa da Escola promove o uso de ferramentas modernas de análise dentro de tópicos de fronteira estudados pelas colaborações experimentais do CERN-LHC.

**Começar (Iniciar)** 4 de nov. de 2024 08:00  
**Fim** 8 de nov. de 2024 18:00  
America/Sao\_Paulo

SPRACE, NCC-UNESP  
R. Dr. Bento Teobaldo Ferraz, 271 - Várzea da Barra Funda, São Paulo - SP, 01140-070  
[Ir para o mapa](#)

**Gustavo Gil Da Silveira**  
Thiago Tomei Fernandez

Ainda não há materiais

<https://indico.cern.ch/e/1escolainct>

Data:  
04 a 08 de novembro de 2024

Localização:  
Núcleo de Computação  
Científica da Unesp  
(NCC-Unesp)

Anúncio programado  
para **05 de julho**



# Campus Unesp – São Paulo

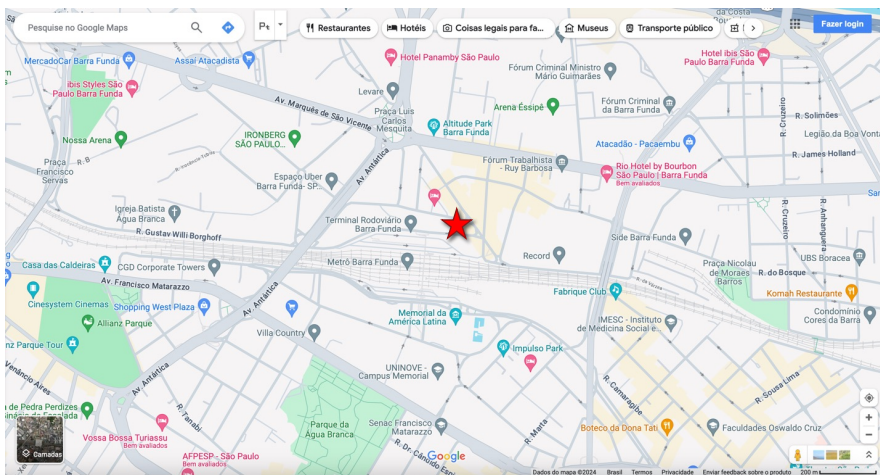


Adjacente à estação intermodal  
Palmeiras-Barra Funda

- ❑ Metrô Butantã (USP): 30 min
- ❑ Trem St. André (UFABC): 35 min
- ❑ Aeroporto Guarulhos: 40 min
- ❑ Campinas (ônibus): 90 min

Hotéis próximos

- ❑ Rio Hotel by Bourbon
- ❑ Hotel Panamby
- ❑ Ibis Styles
- ❑ Hostels variados



# Núcleo de Computação Científica da Unesp











# Ferramentas para a Escola

Baseado no Google Colab para evitar contas CERN

- ❑ Permitir um ambiente de análise simplificado similar ao CERN SWAN

Dados do CERN OpenData ou obtidos com geradores/Delphes

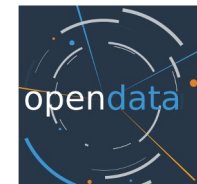
- ❑ Permitir flexibilidade para preparação das atividades de hands-on
- ❑ Evitar acesso a diferentes datasets das Colaborações

ROOT RDataFrame, RooFit, etc.

- ❑ Ferramentas modernas de análise em consonância com os desenvolvimentos atuais nas Colaborações
- ❑ Viabilizar aos estudantes iniciar rapidamente uma análise de baixa ou alta complexidade

GitHub, Docker, etc.

- ❑ Preservação do código e do ambiente de análises necessário para reprodutibilidade e reinterpretação





## INCT CERN Brasil

2 followers Brazil

Follow

### Popular repositories

[Escola2024](#) Public

View as: Public

You are viewing the README and pinned repositories as a public user.

You can [create a README file](#) visible to anyone.

### Repositories

Find a repository... Type Sort New

[Escola2024](#) Public

☆ 0 🍴 0 🕒 0 📄 0 Updated on May 29

### People



# Tópicos da Escola (1)

## Dia #1: ELIZA MELO (UERJ) – Básico de Análise

- ❑ Subsistemas de um detector de partículas
- ❑ Estrutura de dados e variáveis base para cinemática de partículas
- ❑ Ferramentas para extração e ordenamento de dados e formas de análise

## Dia #2: ALEXANDRE SUAIDE (USP) – Básico de Estatística

- ❑ Conceitos básicos de estatística e medidas descritivas
- ❑ Funções de densidade de probabilidade, incertezas, covariância e correlação
- ❑ Métodos de inferência estatística, intervalos de confiança e testes de hipóteses
- ❑ Método de Monte Carlo, dados multivariados, componentes principais (PCA) e linear discrimination analysis (LDA)

# Tópicos da Escola (2)

## Dia #3: EDMAR DE SOUZA (UFBA)– Básico de Aprendizado de máquina

- ❑ Introdução, preparação de conjuntos de dados e sistemas de classificação e regressão
- ❑ Redes neurais artificiais e convolucionais
- ❑ Máquinas de vetores de suporte e técnicas de agrupamento e clusterização

## Dia #4: MARISILVIA DONADELLI (UERJ) – Reinterpretação

- ❑ RECAST – estendendo o impacto de análises existentes
- ❑ Introdução às ferramentas do Docker e GitLab
- ❑ Formato de dados e HepData

## Dia #5: CBPF – Scaling de Análise/Plataforma IaaS

- ❑ Disponibilização de máquinas virtuais e *facilities* de análise

# Programa da Escola

	ELIZA	ALEXANDRE	EDMAR	MARISILVIA	CBPF
	SEG	TER	QUA	QUI	SEX
9:00	Básico de Análise	Básico de Estat.	Básico de ML	Reinterpretação	Scaling de Análise
10:30	BREAK	BREAK	BREAK	BREAK	BREAK
11:00	Análise Avançada	Estatística em FAE	ML em FAE	Reinterpretação	Scaling de Análise
12:30	ALMOÇO	ALMOÇO	ALMOÇO	ALMOÇO	ALMOÇO
14:00	Hands-On	Hands-On	Hands-On	Hands-On	Hands-On
16:00	BREAK	BREAK	BREAK	BREAK	BREAK
16:30	Hands-On	Hands-On	Hands-On	Hands-On	Hands-On
18:30	FECHAMENTO	FECHAMENTO	FECHAMENTO	FECHAMENTO	FECHAMENTO

# Desembolso do GT7 (I Workshop)

Item	Valor Unitário	Quantidade	Valor Total	Detalhamento
<b>Ano 3</b>				
Diárias	R\$ 320,00	150	R\$ 48.000,00	Diárias (5) para 30 alunos para 1a edição da Escola
Passagens	R\$ 1.500,00	30	R\$ 45.000,00	Passagens para 30 alunos para 1a edição da Escola
Diárias	R\$ 320,00	25	R\$ 8.000,00	Diárias (5) para 5 palestrantes para 1a edição da Escola
Passagens	R\$ 1.500,00	5	R\$ 7.500,00	Passagens para 5 palestrantes para 1a edição da Escola
<b>TOTAL ANO</b>			<b>R\$ 108.500,00</b>	Total 1a Edição da Escola
<b>Ano 4 ou 5</b>				
Diárias	R\$ 320,00	150	R\$ 48.000,00	Diárias (5) para 30 alunos para 2a edição da Escola
Passagens	R\$ 1.500,00	30	R\$ 45.000,00	Passagens para 30 alunos para 2a edição da Escola
Diárias	R\$ 320,00	25	R\$ 8.000,00	Diárias (5) para 5 palestrantes para 2a edição da Escola
Passagens	R\$ 1.500,00	5	R\$ 7.500,00	Passagens para 5 palestrantes para 2a edição da Escola
<b>TOTAL ANO</b>			<b>R\$ 108.500,00</b>	Total 2a Edição da Escola
<b>TOTAL</b>			<b>R\$ 217.000,00</b>	

<https://indico.cern.ch/event/1310815/contributions/5537049/>



# Estimativa de Desembolso da 1ª Escola

Dispêndio	Valor Unitário	Quantidade	Valor Total	Detalhamento
Diárias Lecturers	R\$ 320,00	30	R\$ 9.600,00	Diárias (6) para 5 lecturers
Passagens	R\$ 1.000,00	5	R\$ 5.000,00	Passagens para 5 lecturers
Diárias Estudantes	R\$ 320,00	190	R\$ 51.200,00	Diárias (5) para 40 estudantes
Passagens	R\$ 1.000,00	32	R\$ 32.000,00	Passagens para 40 estudantes
Coffee break	R\$ 1.000,00	10	R\$ 10.000,00	Um sessão por turno para 5 dias
<b>TOTAL</b>			<b>R\$ 107.800,00</b>	

## Observações

- ❑ Custo médio da passagem: R\$ 1.000,00
- ❑ Diária CNPq: R\$ 320,00
  - Idêntico para estudantes e doutores
- ❑ Valores a serem otimizados para dar suporte ao maior número de estudantes quanto possível
  - Critérios de análise das inscrições
  - Considerar limitação geográfica em termos de custo das passagens
- ❑ Coffee break: R\$ 1.000,00 / break x 2 turnos/dia x 5 dias = R\$ 10.000,00

# Cronograma da Escola

Anúncio da Escola e lançamento do site: 05 de julho

Início das inscrições: 05 de julho

Final das inscrições: 12 de agosto

Anúncio dos escolhidos: 02 de setembro

Compra de passagens / reserva de hotéis: setembro—outubro

Realização da Escola: **04 a 08 de novembro**

# Próximos Passos

## Reuniões com os lecturers

- ❑ Aperfeiçoamento do conteúdo e das ferramentas computacionais

## Período de seleção

- ❑ Seleção conjunta com todos os membros do GT7
- ❑ Critérios de escolha
  - Colaboração experimental
  - Distribuição geográfica
  - Equidade
  - Excelência