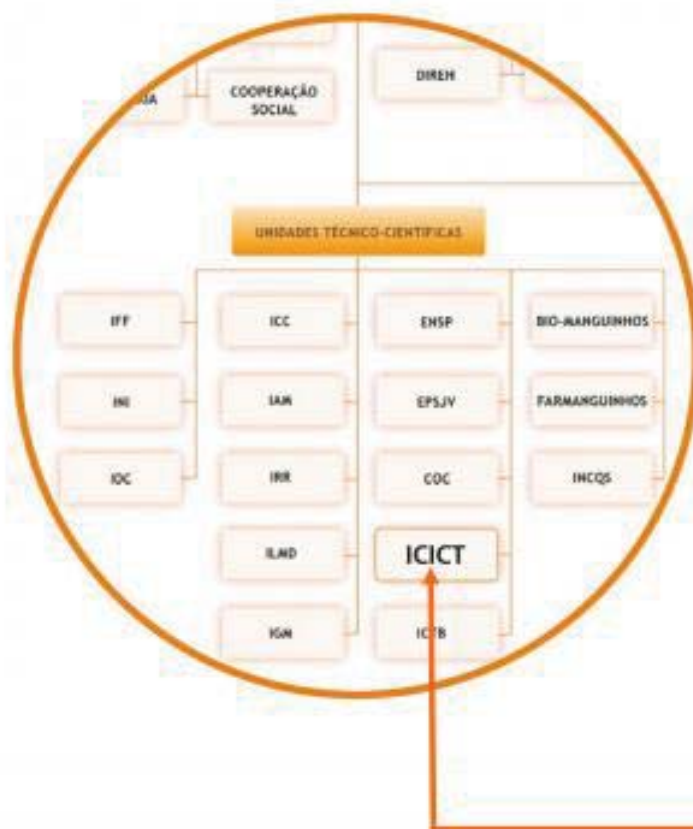


Plataforma de Ciência de Dados aplicada à Saúde

Prof. Dr. Marcel Pedroso
Pesquisador em Saúde Pública

POSIÇÃO DO ICICT NO ORGANOGRAMA DA FIOCRUZ





• O QUE FAZEMOS

- Desenvolver sistemas de informação, aplicativos eletrônicos e plataformas tecnológicas para a Fiocruz e instituições ligadas ao Ministério da Saúde do Brasil e de outros países;
- Preservar e disseminar o acervo bibliográfico e audiovisual da Fiocruz;
- Gerar, analisar e divulgar, em livre acesso para toda a sociedade, informações epidemiológicas, socioeconômicas e ambientais e
- Fomentar a pesquisa e o ensino no campo da comunicação e da informação científica e tecnológica em saúde.

COMUNICAÇÃO CIENTÍFICA E TECNOLÓGICA PARA O SUS

Portais:



• Portal Fiocruz – www.fiocruz.br



• Portal Proqualis – proqualis.net



• Saúde Amanhã – saudeamanha.fiocruz.br



• PenseSUS – pensesus.fiocruz.br

INFORMAÇÃO CIENTÍFICA E TECNOLÓGICA PARA O SUS

Sistemas:

- **Sistema Nacional de Informações Tóxico-Farmacológicas (Sinitox)** - www.fiocruz.br/sinitox



- **Projeto de Avaliação de Desempenho do Sistema de Saúde (ProAdess)** - www.proadess.icict.fiocruz.br



- **Sistema de Indicadores de Saúde e Acompanhamento de Políticas do Idoso (Sisap Idoso)** - sisapidoso.icict.fiocruz.br



INFORMAÇÃO CIENTÍFICA E TECNOLÓGICA PARA O SUS

Repositórios:



- Registro Brasileiro de Ensaios Clínicos (Rebec)

www.ensaiosclinicos.gov.br



- Repositório Institucional da Fiocruz (Arca)

www.arca.fiocruz.br

ENSINO PARA O SUS

– *Stricto Sensu:*

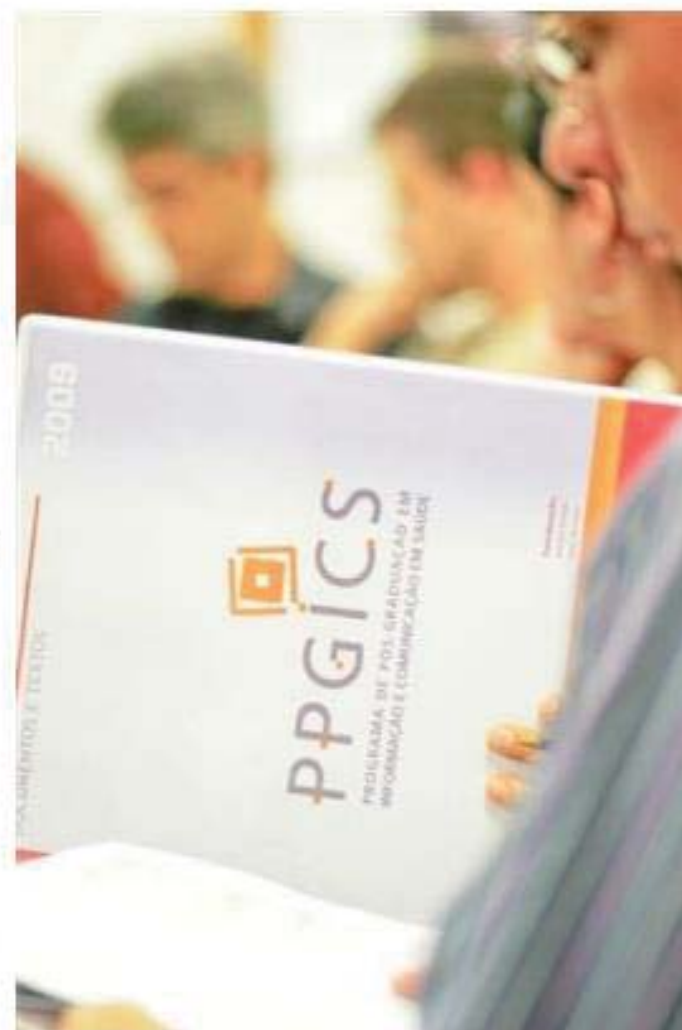
- Mestrado acadêmico em Informação e Comunicação em Saúde
- Doutorado em Informação e Comunicação em Saúde

– *Lato Sensu:*

- Especialização em Comunicação e Saúde
- Especialização em Informação Científica e Tecnológica em Saúde

– **Cursos de Atualização e Extensão:**

- Presencial e a distância



PESQUISA CIENTÍFICA PARA O SUS

Laboratórios de Pesquisa:



- Laboratório de Informação em Saúde (LIS)



- Laboratório de Informação Científica e Tecnológica em Saúde (LICTS)



- Laboratório de Comunicação e Saúde (LACES)

CIÊNCIA DE DADOS - DEFINIÇÃO

Ciência de Dados é um conjunto de estratégias, ferramentas e técnicas que busca reunir equipes multidisciplinares formadas por pesquisadores com conhecimento substantivo do problema em análise - no nosso caso saúde pública - estatísticos, matemáticos e cientistas da computação. Trata-se de um campo de estudo promissor e destaca-se pela capacidade de auxiliar a descoberta de informação útil a partir de grandes bases de dados e a tomada de decisão orientada por dados



OBJETIVO PRINCIPAL

Pesquisar, desenvolver e disponibilizar Plataforma de Ciência de Dados para armazenamento, gestão e análise de Big Data em Saúde para **pesquisadores, docentes e discentes de instituições de ensino e pesquisa, bem como, **gestores governamentais****

OBETIVOS ESPECÍFICOS

- ✓ Pesquisar e avaliar **ferramentas tecnológicas e estratégias** para o desenvolvimento da Plataforma
- ✓ Identificar, coletar, normalizar, armazenar e conectar as **bases de dados de interesse**
- ✓ Adquirir **equipamentos**, desenvolver, testar e homologar a Plataforma
- ✓ Construir **ferramentas e aplicativos** para análise visual, análise multicritério, mineração de dados e análise preditiva
- ✓ Capacitar **equipe de referência** em Ciência de Dados

INSTITUIÇÕES PARCEIRAS



Laboratório
Nacional de
Computação
Científica



DEXL LAB
EXTREME DATA LAB

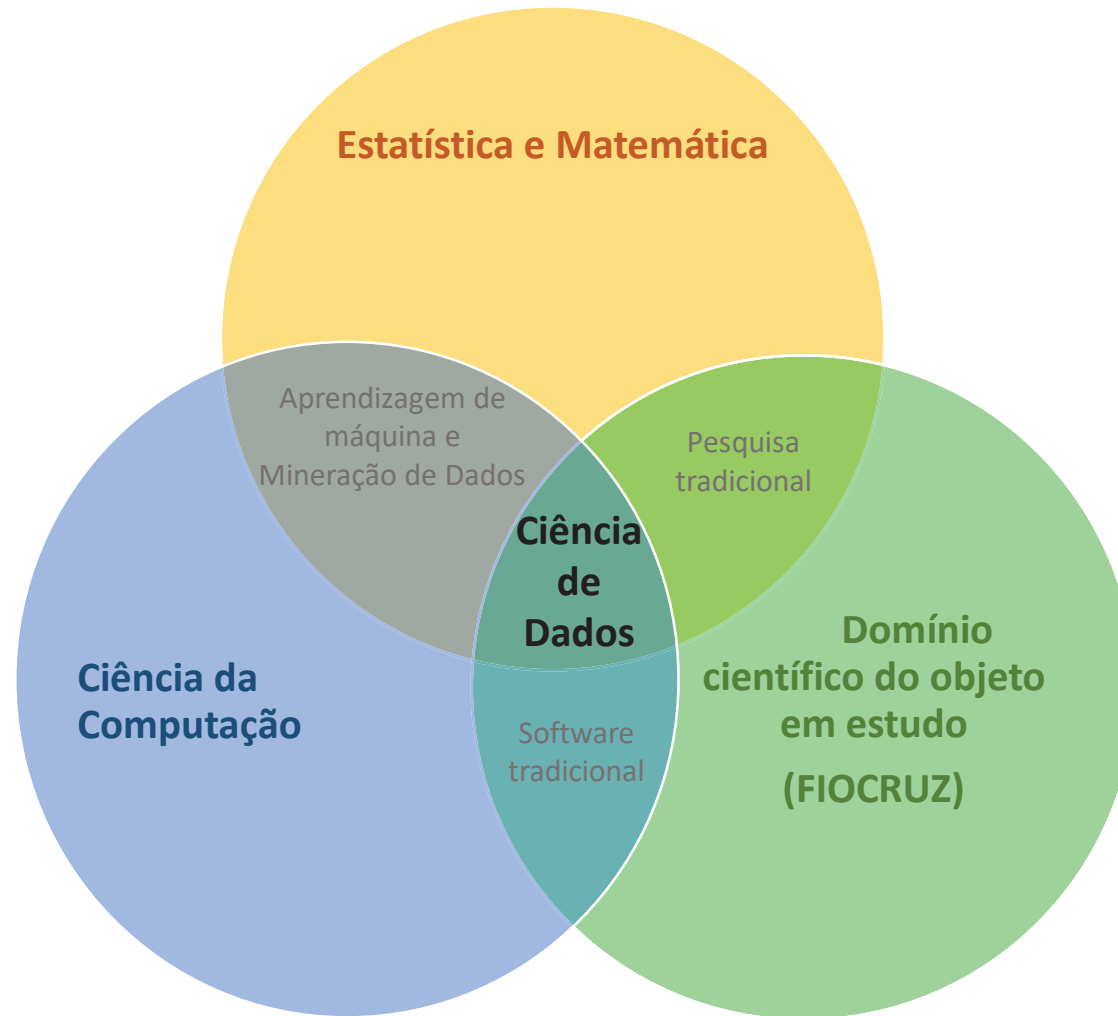


EIC

Escola de Informática e Computação



Equipe de Referência em CIÊNCIA DE DADOS

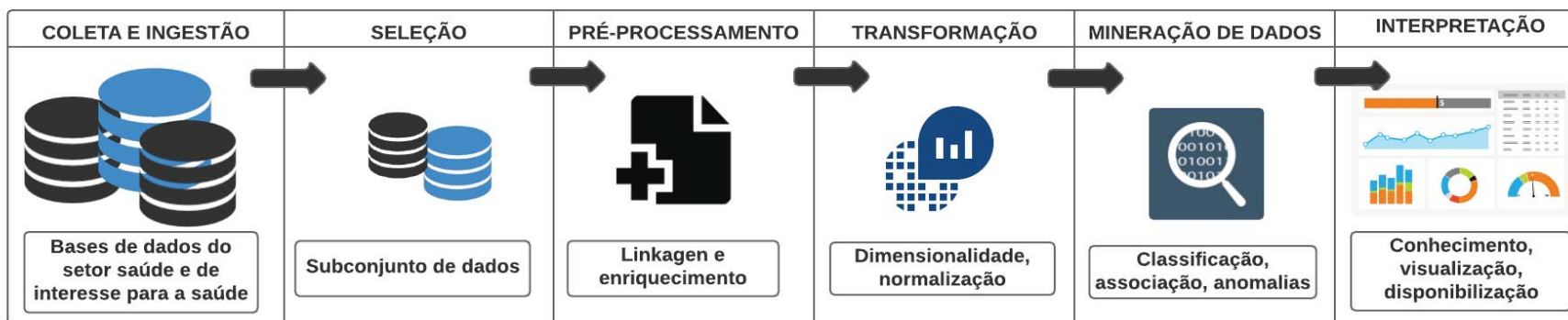


VOLUME E VARIEDADE (até 2015)

- ✓ **180 milhões** de autorizações de internações hospitalares (Datapus)
- ✓ **57 milhões** de registros de nascimentos (Datapus)
- ✓ **19 milhões** de declarações de óbito (Datapus)
- ✓ **24 bilhões** de registros de procedimentos ambulatoriais (Datapus)
- ✓ **2,5 bilhões** de registros de doses de vacinas (Datapus)
- ✓ **120 milhões** de questionários domiciliares dos Censos de 2000 e 2010 (IBGE)
- ✓ **milhões** de medições diárias realizadas pelo INPE e CEMADEN

- ✓ diariamente são criados **2,5 quintilhões** de dados, sendo que **3/4** deste conteúdo é produzido por internautas (dados interacionais) e não empresas ou instituições (dados transacionais)
- ✓ As **interações** com Google geram diariamente 24 petabytes (1 PB é o suficiente para armazenar, por exemplo, um vídeo em alta definição com duração de 13,3 anos)
- ✓ Os 1.3 bilhões de **usuários** do Facebook (68 milhões no Brasil) compartilham mais de 350 milhões de fotos e geram 500 terabytes por dia

Descoberta de Conhecimento em Bases de Dados Knowledge Discovery in Databases – KDD



PLATAFORMA DE CIÊNCIA DE DADOS como serviço (PaaS)



COMPONENTES DA PLATAFORMA

Indexação, Extração e Análise Visual



Mineração de Dados e Análise Preditiva



Inovação e Aprendizagem Colaborativa



COMPONENTE ANÁLISE VISUAL

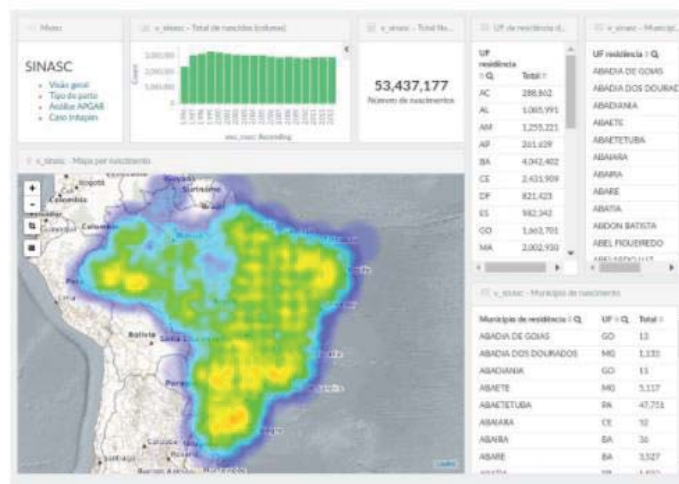


Análise Visual

Indexação, extração e análise visual de grandes quantidades de dados do setor saúde e seus determinantes socioambientais

Acesso qualificado aos dados da Plataforma

- Indexação e disponibilização de grandes bases de dados
- Extração de subconjuntos de dados de interesse para os pesquisadores
- Análise visual de situações de saúde
- Processamento distribuído e escalável



[Acesse a plataforma](#)

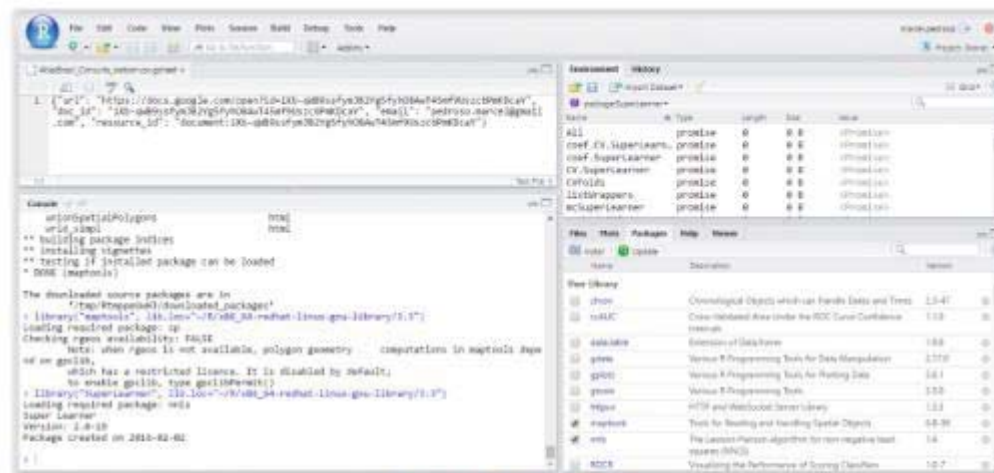
COMPONENTE MINERAÇÃO DE DADOS



Mineração de Dados e Análise Preditiva Conexão aos dados da Plataforma via R Studio Server

Conexão e análise dos dados na Plataforma utilizando os pacotes R

- Conexão dos pacotes R sem necessidade de movimentação dos dados
- Utilização da infraestrutura de processamento e armazenamento da Plataforma
- Processamento dos dados por meio de escalonamento, replicação e indexação via Elasticsearch



Acesse a plataforma

COMPONENTE DATA SCIENCE LAB



Data Science Lab Inovação e Aprendizagem Colaborativa

<p>Kaggle Plataforma de aprendizagem e competição para cientistas de dados.</p>	<p>Jupyter Python Plataforma de ciência de dados para compartilhamento de códigos em mais de 40 linguagens de programação.</p>	<p>Curso online de Ciência de Dados Programa de cursos da Universidade Johns Hopkins na Coursera.</p>	<p>Curso online de Ciência de Dados Aprenda a programar em Python de um modo profissional e divertido.</p>
<p>Big Data University - IBM Cursos gratuitos em português sobre Big Data.</p>	<p>GitHub Educação Codificação - Cursos e Ferramentas grátis para estudantes.</p>	<p>Codecademy Aprenda a programar de maneira interativa e gratuita.</p>	<p>Social Data Analytics Mineração de textos e análise de sentimento por meio de redes sociais pela equipe IBM Research - Brasil.</p>
<p>DataCamp Aprenda Python e R na prática em plataforma inovadora de aprendizagem ciência de dados e machine learning.</p>	<p>HARVARD UNIVERSITY The Harvard Data Science Initiative Programa de pesquisa e ensino interdisciplinar que reuniu esforços em toda a Universidade de Harvard para o fomento de parcerias visando a fomento de iniciativas em Ciência de Dados.</p>		



BRASIL Acesso à informação Participe Serviços Legislação Canais

FIOCRUZ Fale com a Fiocruz FUNDAÇÃO OSWALDO CRUZ

ICICT Instituto de Comunicação e Informação Científica e Tecnológica em Saúde

Ciência de Dados aplicada à Saúde

Buscar no site

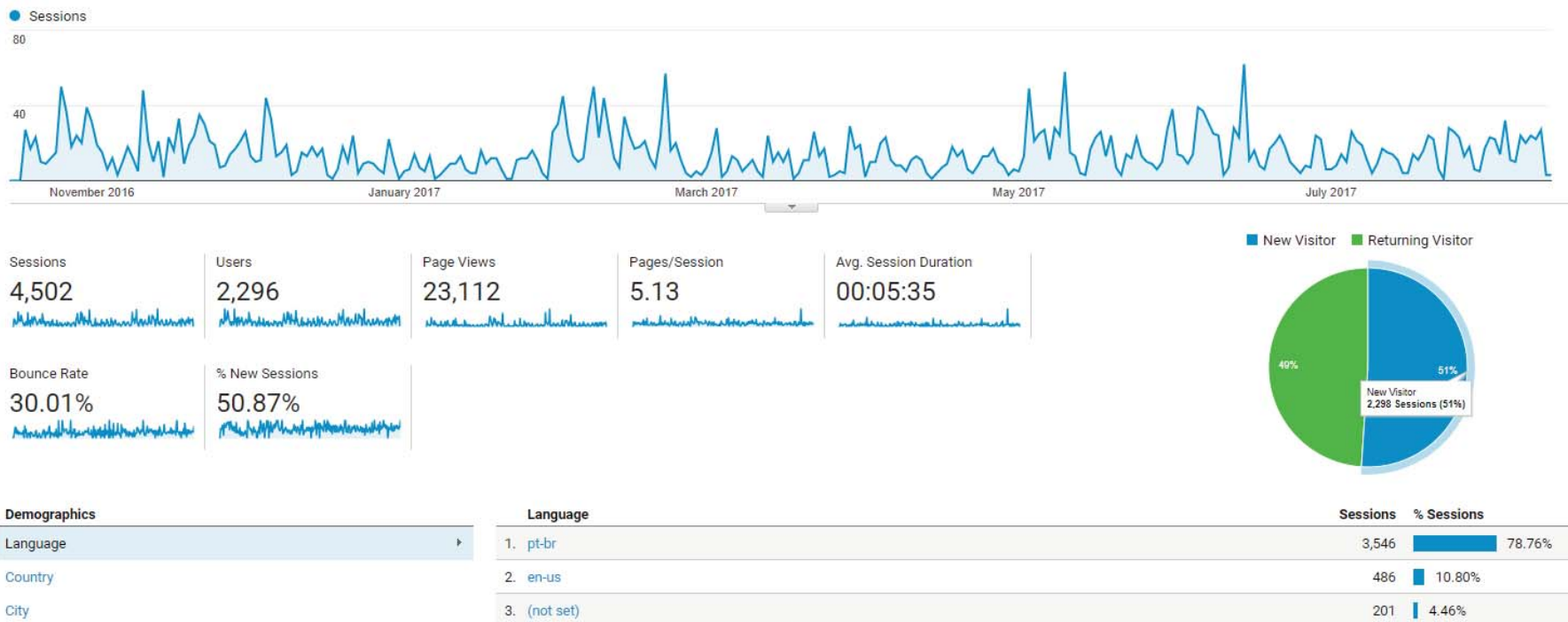
INTERFACE TECNOLÓGICA GALERIA VISUAL DE DADOS ACESSE A PLATAFORMA CAPACITAÇÃO EQUIPE INSTITUIÇÕES PARCEIRAS CONTATO

Ciência de Dados é um conjunto de estratégias, ferramentas e técnicas que busca reunir equipes multidisciplinares formadas por pesquisadores com conhecimento substantivo do problema em análise - no nosso caso saúde pública - estatísticos, matemáticos e cientistas da computação. Trata-se de um campo de estudo promissor e destaca-se pela capacidade de auxiliar a descoberta de informação útil a partir de grandes bases de dados e a tomada de decisão orientada por dados



LEIA MAIS

INFORMAÇÕES SOBRE UTILIZAÇÃO DA PLATAFORMA



2.200+ usuários diferentes acessaram o site

300+ usuários cadastrados na Plataforma

5 minutos em média de tempo de visita

51% de novos acessos

HARDWARE

- **2 servidores para de gestão do cluster**
- 4 Processadores Intel Xeon E5-2630 v3 com 8 Núcleos e 2.4GHz cada
- 256 GB memória RDIMM total (126 GB por servidor)
- **4 servidores para armazenamento de dados**
- 40 Terabyte de armazenamento
- 8 Processadores Intel Xeon E5-2630 v3 com 8 Núcleos e 2.4GHz cada
- 64 GB memória RDIMM total (16 GB por servidor)
- **10 Gigabit conexão internet**



CAPACITAÇÃO em CIÊNCIA DE DADOS

✓ EQUIPE DE REFERÊNCIA

ICICT - LNCC - USP – CEFET - MS - UnB

Plataforma online de capacitação em Ciência de Dados **DataCamp**



✓ DISCIPLINA e ATUALIZAÇÃO - ICICT

Ciência de Dados aplicada à Saúde

Curso presencial (120 horas) TCC utilizando a Plataforma

CAPACITAÇÃO em CIÊNCIA DE DADOS

Participação de integrantes do Icict no 13º Programa de Verão do LNCC, em Petrópolis.



Fonte: Matéria da Ascom/Icict intitulada "Grupo de Pesquisa da Fiocruz adota estratégia Big Data em Saúde".

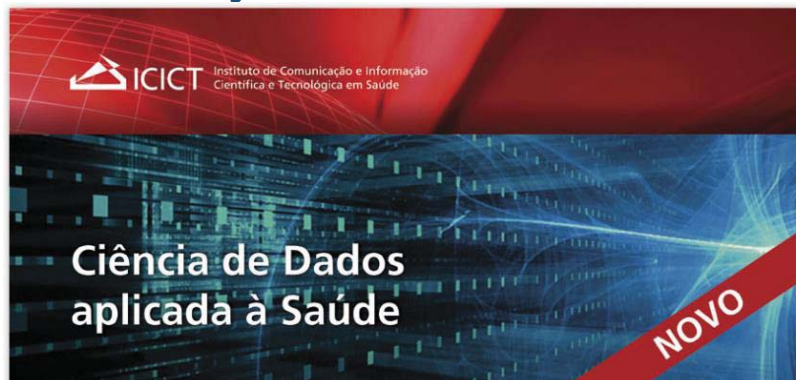
CAPACITAÇÃO em CIÊNCIA DE DADOS

Seminário para lançamento da Plataforma de Ciência de Dados aplicada à Saúde



Fonte: Matéria da Ascom/Icict intitulada “Especialistas em Ciência de Dados e Saúde discutem ferramentas e técnicas em Seminário do Centro de Estudos”.

CAPACITAÇÃO em CIÊNCIA DE DADOS



CURSO DE ATUALIZAÇÃO - 2017

Coordenação: Marcel Pedroso e Christovam Barcellos

Carga Horária: 60 horas

Local: Rio de Janeiro/RJ

10 VAGAS

Período do Curso:

22 de março a 12 de julho de 2017 (somente às 4^{as} feiras)
das 14h às 17h30min.

Inscrições:

02 de janeiro a 02 de fevereiro de 2017

Plataforma SIGA – www.sigals.fiocruz.br

Seguindo os links: Inscrição>Presencial>Atualização>Icict>Ciência de Dados aplicada à Saúde – 2017/Sede

Atenção: Antes de se inscrever, leia atentamente a chamada de Seleção Pública 2017.

Gestão Acadêmica: Av. Brasil 4036 - Prédio da Expansão do Campus - sala 210 - Manguinhos - RJ - Cep: 21040-361
Tel.: (55xx21) 3882-9063/9033 • e-mail: gestaoademica@icict.fiocruz.br

CAPACITAÇÃO em CIÊNCIA DE DADOS

Aula e visita técnica da 1ª turma de Ciência de Dados aplicada à Saúde ao LNCC



Fonte: Matéria da Ascom/Icict intitulada Alunos do curso de Ciência de Dados visitam o LNCC, em Petrópolis”.

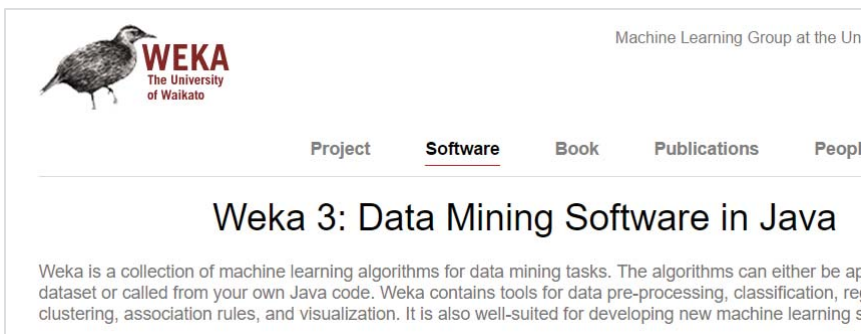
PÓXIMOS PASSOS - MINERAÇÃO DE DADOS E MACHINE LEARNING VISUAL



H₂O.ai Products Industries Customers Community

Why H2O? H2O Flow H2O on R H2O on Python Machine Learning

Better Machine Learning, End-to-End

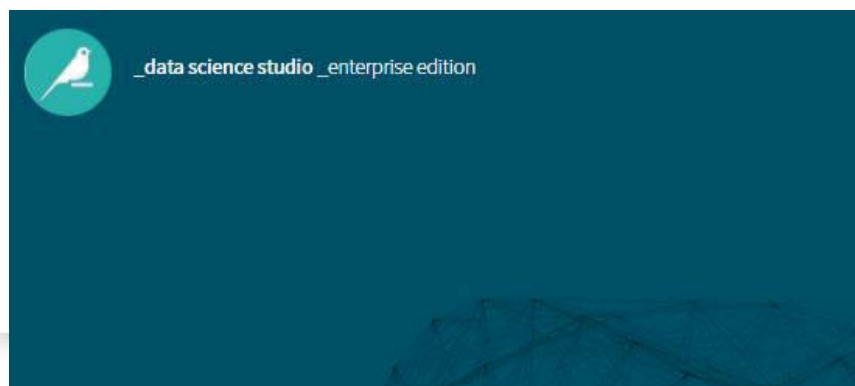


WEKA The University of Waikato Machine Learning Group at the University of Waikato

Project **Software** Book Publications People

Weka 3: Data Mining Software in Java

Weka is a collection of machine learning algorithms for data mining tasks. The algorithms can either be applied to a dataset or called from your own Java code. Weka contains tools for data pre-processing, classification, regression, clustering, association rules, and visualization. It is also well-suited for developing new machine learning software.



data science studio enterprise edition

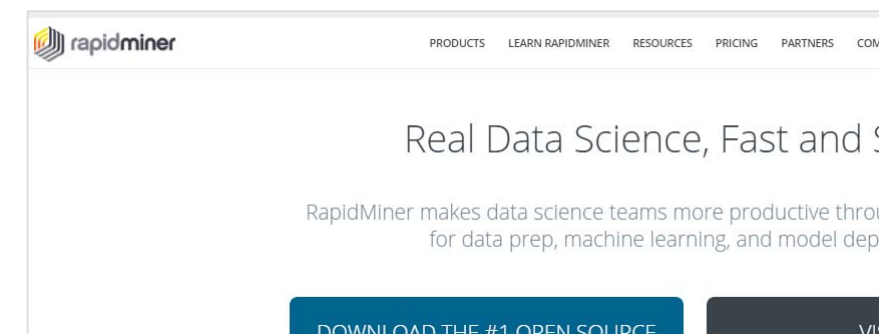


Open for Innovation **KNIME** PRODUCTS / SOLUTIONS / LEARNING / PARTNERS / COMMUNITY / ABOUT

Open for Innovation

Navigate complex data with the agility and freedom that only an open platform can bring

[Learn More](#)

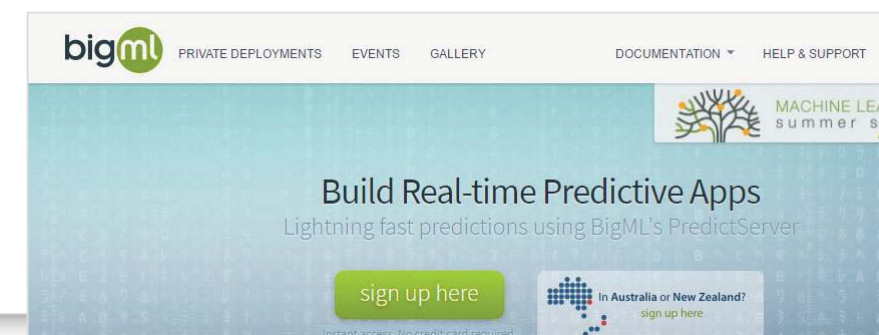


rapidminer PRODUCTS LEARN RAPIDMINER RESOURCES PRICING PARTNERS COMMUNITY

Real Data Science, Fast and Simple

RapidMiner makes data science teams more productive through automation for data prep, machine learning, and model deployment.

[DOWNLOAD THE #1 OPEN SOURCE](#)



bigml PRIVATE DEPLOYMENTS EVENTS GALLERY DOCUMENTATION HELP & SUPPORT

Build Real-time Predictive Apps

Lightning fast predictions using BigML's PredictServer

[sign up here](#)

Instant access. No credit card required.

In Australia or New Zealand? [sign up here](#)

GRUPO DE PESQUISA NO CNPq



The screenshot shows the CNPq website interface for a research group. The header includes the CNPq logo and the 'Diretório dos Grupos de Pesquisa no Brasil Lattes' logo. The main content area displays the group name 'Ciência de Dados aplicada à Saúde' and its address. A left sidebar contains a menu with options like 'Identificação', 'Endereço / Contato', and 'Repercussões'. The main content area shows the group's identification details, including its status as 'Certificado', formation year (2014), and the names of its leaders: Christovam Barcellos and Marcel de Moraes Pedrosa. A 'CERTIFICADO PELA INSTITUIÇÃO' seal is also visible.

Identificação

Endereço / Contato

Repercussões

Linhas de pesquisa

Recursos humanos

Instituições parceiras

Indicadores de RH

Equipamentos e Softwares

Grupo de pesquisa

Ciência de Dados aplicada à Saúde

Endereço para acessar este espelho: dgp.cnpq.br/dgp/espelhogrupo/4230691756969719

Identificação

Situação do grupo: Certificado

Ano de formação: 2014

Data da Situação: 11/12/2014 09:53

Data do último envio: 18/05/2016 14:36

Lider(es) do grupo: Christovam Barcellos

Marcel de Moraes Pedrosa

Área predominante: Ciências da Saúde; Saúde Coletiva

Instituição do grupo: Fundação Oswaldo Cruz - FIOCRUZ

Unidade: Centro de Informação Científica e Tecnológica

<http://dgp.cnpq.br/dgp/espelhogrupo/4230691756969719>

OUTRAS PARCERIAS? (Instituições e pesquisadores)

marcel.pedroso@icict.fiocruz.br

CRONOGRAMA

Meta	Fase	mês 1	mês 2	mês 3	mês 4	mês 5	mês 6	mês 7	mês 8	mês 9	mês 10	mês 11	mês 12	mês 13	mês 14	mês 15	mês 16	mês 17	mês 18	mês 19	mês 20	mês 21	mês 22	mês 23	mês 24
1	FONTES DE DADOS																								
1.1	Pesquisar e avaliar ferramentas tecnológicas e estratégias para o desenvolvimento da infraestrutura																								
1.2	Identificar, coletar, normalizar, armazenar e conectar as bases de dados de interesse																								
2	GESTÃO E ARMAZENAMENTO																								
2.1	Adquirir equipamentos para a Plataforma																								
2.2	Desenvolver, testar e homologar a Plataforma																								
3	ANÁLISE E VISUALIZAÇÃO																								
3.1	Análise visual, análise multicritério, mineração de dados e análise preditiva de Big Data em Saúde																								
3.2	Capacitar equipe de referência em Computação Científica e Big Data																								

PRINCIPAIS PARCEIROS





Institute for Scientific and Technological Communication and Information on Health

www.facebook.com/fiocruz.icict

twitter.com/@Icict_fiocruz

www.youtube.com/videosaudefio

www.icict.fiocruz.br