

BOLETIM INFORMATIVO DO SER

Volume 2/ Edição 1/Junho-2017 – Luciane F. Alcoforado

II SER – MAIS UMA EDIÇÃO DE SUCESSO

A segunda edição do Seminário Internacional de Estatística com R nos mostra a força desta poderosa linguagem denominada R, seja através de grande participação do público como também dos pesquisadores que enviaram seus trabalhos ou ministraram palestras, oficinas e minicursos.

Renovou-se a missão em promover um intercâmbio de conhecimento entre pesquisadores e usuários da linguagem R, promovendo o encontro entre diversas áreas do conhecimento.

A primeira edição marcou o pioneirismo do evento no Brasil, inserindo-se no calendário anual dos pesquisadores acadêmicos e profissionais de análises de dados.

SER abre o primeiro dia com os R-pesquisadores da Espanha, Portugal e do Brasil

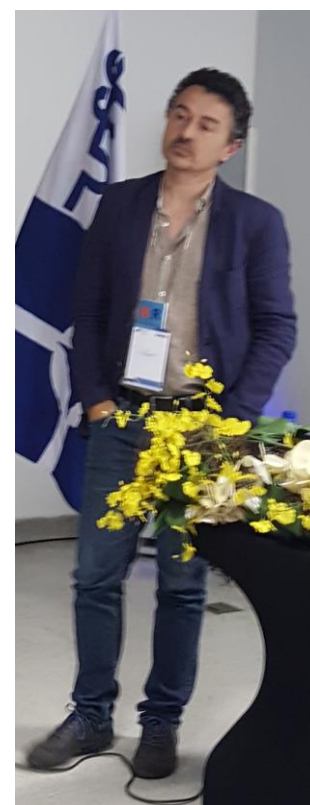
O II SER – Seminário Internacional de Estatística com R ocorreu nos dias 23 e 24 de maio de 2017 no auditório do Núcleo de Águas e Biomassas (NAB) da UFF, contando com a participação de alunos de Graduação, Pós-Graduação, Professores e Profissionais do mercado, interessados no aprimoramento e nas aplicabilidades da linguagem R.

Abrindo o evento com o espaço blog, **Sillas Gonzaga** se apresenta pela segunda vez mostrando sua paixão por dados e pela linguagem R. Seu primeiro contato com o R foi em intercâmbio na Coréia do Sul quando ainda era estudante de graduação. Criou o **blog Paixão por Dados** em que mostra diversas aplicações produzidas por ele utilizando o R. Seu blog hoje conta com mais de 3 mil acessos e mais de 10 mil visualizações. Ele já possui um pacote no cran do R chamado **mafs** para ajuste de modelos preditivos de séries temporais.

Abrindo o espaço de conferências tivemos a presença do **prof. Luís Torgo** da Universidade do Porto em Portugal. O prof. Luís utiliza o R desde o final da década de 90, programando pacotes, códigos no R e conta com uma importante produção sobre Data Mining.



Em sua conferência ele expos as funcionalidades do pacote criado por ele, o **Performance Estimation** que trata sobre facilitar a seleção de modelos preditivos: **“o pacote vai tentar enfrentar este desafio: qualquer que seja a tarefa de previsão que tenhamos, há de fato um conjunto enorme de possíveis técnicas que podemos utilizar e dentro de cada técnica um conjunto variável de parâmetros que podem ser ajustados e uma centena de abordagens que podemos aplicar para resolver o modelo de previsão concreto”**, explicou o prof. Torgo.



Prof. Luís Torgo em sua conferência no II SER.
Foto: Equipe SER

REPRESENTATIVIDADE DE NO II SER

O evento contou com a participação de representantes de 15 estados brasileiros e 3 países da Europa: Espanha, Itália e Portugal.

Entre os participantes inscritos haviam alunos de graduação e pós-graduação, professores de diversas instituições de ensino como UFF, UNIRIO, ENCE, FGV, CESGRANRIO, CEFET/RJ, UDESC, IFFluminense, UNIVERSO, PUC/RJ, UFCG, UERJ, UFBA, UFRGS e UNIMONTES; profissionais do mercado de diversos setores como Ancine, INEP, Banco do Brasil, Ministério do Comércio Exterior, SEEDUC, Fundação CECIERJ dentre outros.

Pela segunda vez no evento, os conferencistas **prof. Manuel Febrero Bande** da Universidade de Santiago de Compostela da Espanha e o **prof. Pedro Ferreira** da Fundação Getúlio Vargas do Brasil, ambos discorrendo sobre o tema Séries Temporais.



O prof. Manuel Febrero abordou os conceitos clássicos de séries temporais e indicou os principais aspectos que devem ser levados em conta pelo pesquisador que utiliza-se desta técnica com o apoio do R. “**Nós estudamos o passado, vemos todas as estruturas do passado que há e predizemos o futuro. Não usamos bola de cristal, usamos matemática**”, disse o prof. Febrero.



O prof. Pedro Ferreira abordou a temática de Séries Temporais explorando as funcionalidades do pacote **BETS** que já encontra-se disponível no CRAN. Foi um projeto que iniciou há 1 ano e meio atrás e apoia a pesquisa aplicada do Brasil, além de dar maior visibilidade as séries produzidas pela FGV. Preenche uma lacuna na obtenção e unificação de dados: “**Conseguimos colocar no pacote BETS séries temporais do Banco Central, do IBGE e da Fundação Getúlio Vargas**”, relatou o prof. Pedro.



Na abertura do evento estavam presentes o prof. Vitor Francisco Ferreira representando o Magnífico Reitor da UFF e a Pró-Reitoria de Pesquisa, Pós-

Graduação e Inovação; o prof. José Rodrigues de Farias Filho pró-reitor da Pró-Reitoria de Graduação; o prof. Celso José da Costa, Diretor do Instituto de Matemática e Estatística; o prof. José André de Moura Brito representando a ENCE e o IBGE; o prof. Jorge Passamani Zubelli representando o IMPA e a SBMAC e a profa. Luciane Ferreira Alcoforado coordenadora geral do II SER.

A mesa redonda sobre a temática “The WoRld of Big Data Analysis” teve como representantes ilustres

o prof. Ariel Levy como moderador, o prof. Orlando Celso Longo coordenador do Programa de Pós-Graduação em Engenharia Civil, o prof. Eduardo Camilo da Silva coordenador do Programa de Pós-



Mesa Redonda sobre “The WoRld of Big Data Analysis
Foto: Equipe SER

Graduação em Administração, o prof. Jorge Passamani Zubelli do IMPA, o prof. Djalma Galvão Carneiro Pessoa atualmente consultor do IBGE, o prof. Pedro Guilherme Ferreira do IBRE/FGV, o prof. Luis Torgo da Universidade do Porto e os professores Wenceslau Gonzales Manteiga e Manuel Febrero Bande da Universidade de Santiago de Compostela.

O prof. Manteiga nos conta que hoje ele divide a Estatística em 3 classes: A Clássica dos anos 70, a Computacional e a do Big Data, esta última cuja metodologia não está consolidada mas que vem aumentando a demanda por consultoria e desenvolvimento de metodologias estatísticas neste campo do conhecimento. Ele constata que um impacto importante desta era do Big Data é a crescente substituição do termo Estatístico por Cientista de Dados, resultante da simbiose entre o termo Big Data e as demandas da indústria. Por fim ele destaca que os Estatísticos deixam de ser apenas metodológicos e passam a trabalhar em colaboração com os profissionais das ciências da computação.



Mesa Redonda sobre “The WoRld of Big Data Analysis
Foto: Equipe SER

O prof. Djalma esclarece que apesar de o IBGE trabalhar com grandes bases de dados, os mesmos não se enquadrariam com dados de big data, pois são dados coletados através de um desenho amostral, o que permitem fazer inferências e estimar a precisão das estimativas. Já os

dados do Big Data não são dados oriundos de desenhos amostrais como os do IBGE, podendo levar a um viés de seleção. Ele finaliza lançando a pergunta: como se pode superar este problema?.

O prof. Zubelli diz que o termo Big Data é um conceito que vem mudando com o tempo, o que não muda são os desafios matemáticos que devem ser superados. Ele destaca que o conhecimento do R não substitui conhecimento Estatístico, ou seja, junto com o aprendizado do R o pesquisador deve formar uma **sólida base em matemática e estatística**.

O prof. Eduardo esclareceu que hoje temos acesso a muita informação, o problema reside em como extrair deste universo o que é relevante. Ele explica que a palavra Big Data vem por conta das tecnologias de engenharia de sistemas que permitiram que estas bases de dados fossem analisadas. **Para a administração o que interessa são os termos Data Science e Analytics com o objetivo de melhorar as tomadas de decisão e a compreensão dos negócios.**

O prof. Pedro diz que o termo Big Data é a revolução que estamos vivendo em todas as áreas. Ele acabou de escrever um artigo sobre a 4ª. revolução industrial e nos aponta que o mundo já vive um sistema de produção integrado com a internet das coisas atuando de tal forma que as indústrias vão saber se um determinado produto que você comprou já está acabando e como produzir



Mesa Redonda sobre “The World of Big Data Analysis, prof. Luís Torgo, Pedro Ferreira e Manuel Febrero.
Foto: Equipe SER

um produto com as características individuais dos usuários. **Isso já está acontecendo e irá gerar muitos dados que teremos que analisar.**

O prof. Febrero nos fala sobre o modismo do termo Big Data visto que o termo Estatística, em geral, não tem uma boa imagem nas empresas que logo pensam numa tabela e não em algo interessante. Assim, conclui que de **tempos em tempos há que se mudar o nome das coisas, primeiro foi Data Mining, depois Machine Learning e agora Big Data.** No entanto em seu entendimento ele diz que estamos enfrentando uma nova fronteira para o que chama de Estatística para dados complexos seja por termos excesso de número de casos ou de variáveis, ou ainda, por termos diversidade no tipo de informação a ser analisada.



O prof. Torgo destacou o ponto sobre a diversidade dos dados, pois as pessoas normalmente se ligam ao volume dos dados que é apenas um dos aspectos. **O grande desafio é buscar os dados em muitas fontes diferentes e em tempo real. Hoje em dia queremos analisar som, imagem, imagens de satélites, vídeos e não apenas dados em tabelas.** Esse tipo de dados que não se encontram só em tabelas e nos lançam por vezes muitos desafios, especialmente no pré-processamento é o que o futuro nos reserva. Ele encerra sua fala concluindo que **o Big Data não diz respeito apenas ao volume de dados mas também a sua diversidade em fontes de diferentes naturezas.**

O primeiro dia de atividades foi marcado por uma plateia participativa que promoveu uma intensa troca de experiências. Os palestrantes se mostraram acessíveis ao público diversificado, contemplando muitas pessoas iniciantes na linguagem R. Sillas, que abriu o evento com a missão de motivar os iniciantes, relatou de forma brilhante sua experiência, seus projetos e a paixão pelos dados.



Sillas Gonzaga apresentando sua palestra no Espaço Blog.
Foto: Equipe SER



Equipe de recepção coordenada pela Elizete e Raquel.
Foto: Equipe SER

Plateia no primeiro dia do evento.
Foto: Equipe SER



O evento contou ainda com um laboratório de experimentação ao final do primeiro dia do evento, em que os participantes podiam experimentar um script para iniciantes, elaborado pela equipe do projeto Estatística é com R! da UFF, ora coordenado pela profa. Luciane Alcoforado.



Laboratório de experimentação.
Foto: Equipe Estatística com R



Recepção aos palestrantes internacionais da Espanha e Portugal.
Foto: Equipe SER

Na recepção aos palestrantes, contamos com a equipe da UNITEVE que realizou diversas entrevistas para o boletim eletrônico da UFF.



O coffee break foi ofertado pela Escola Nacional de Ciências Estatísticas (ENCE/IBGE), que vem apoiando o evento SER desde a sua primeira edição. Profa. Maysa participou ativamente da avaliação dos trabalhos da sessão pôster constatando que a qualidade dos mesmos vem melhorando em cada nova edição do SER.



Profas. Maysa Magalhães (ENCE) e Luciane Alcoforado (UFF) renovando votos da parceria ENCE-UFF no II SER.
Foto: Equipe SER

O segundo dia foi marcado por:

Oficinas e Minicursos

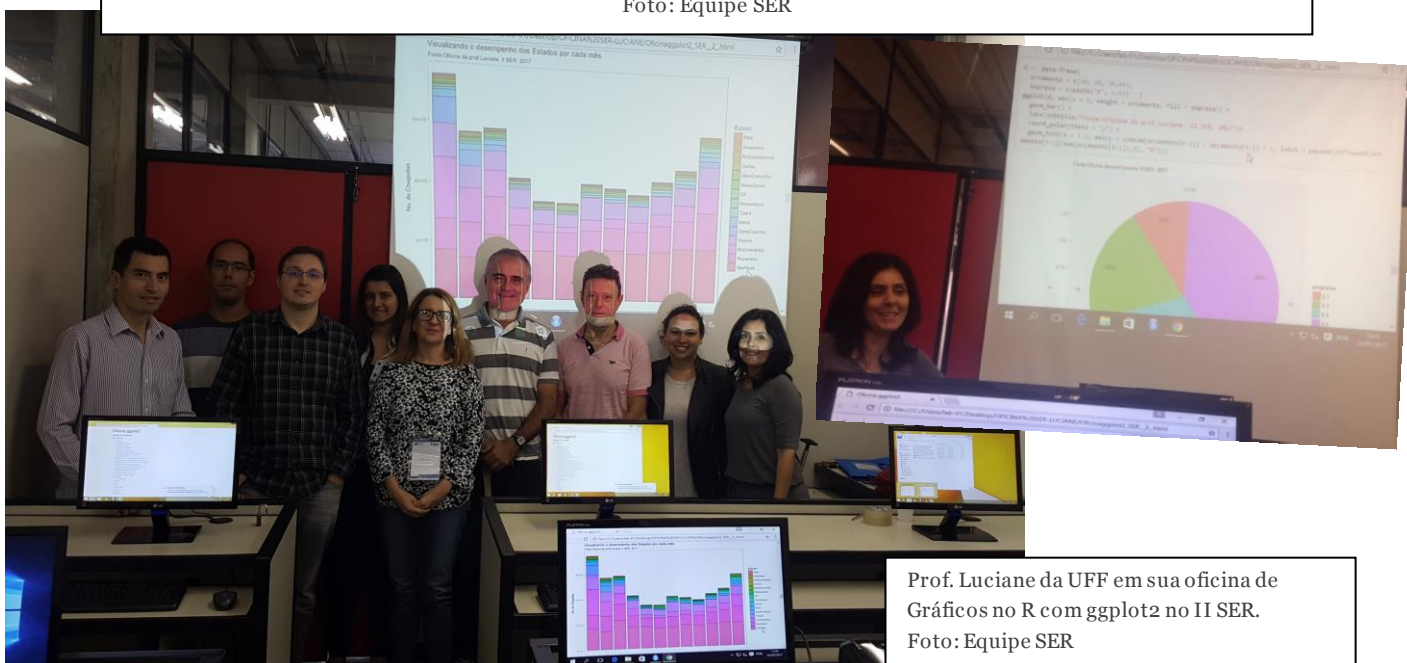
Nesta edição do evento tivemos ampliação da oferta de minicursos e oficinas que abordaram os principais temas do mundo do R. Os participantes puderam escolher entre 10 opções disponíveis, algumas em laboratório de informática como no caso das oficinas e outras em sala de aula como no caso dos minicursos. As atividades foram desenvolvidas simultaneamente por três campus da Universidade: Valonguinho, Gragoatá e Praia Vermelha.



Prof. Pedro Ferreira da FGV em seu minicurso sobre o pacote BETS no II SER.
Foto: Equipe SER



Prof. Alex Laier da UFF em sua oficina sobre Inteligência Artificial com R no II SER.
Foto: Equipe SER



Prof. Luciane da UFF em sua oficina de Gráficos no R com ggplot2 no II SER.
Foto: Equipe SER

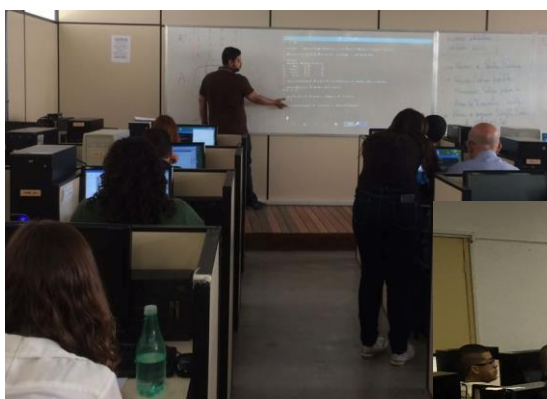


Profa. Ludmilla da UFF em seu minicurso de Estatística Multivariada Aplicada com R no II SER.
Foto: Equipe SER

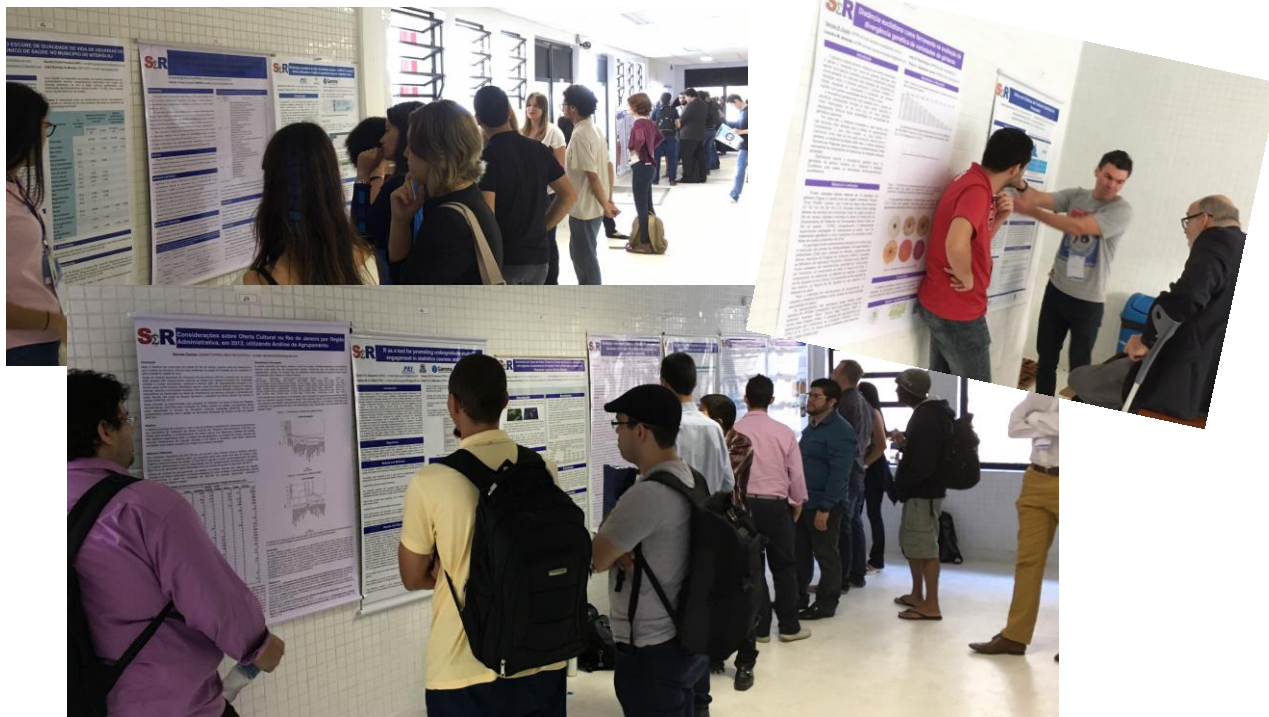


Prof. José Rodrigo da UFF em seu minicurso de Modelos Lineares Generalizados com R no II SER.
Foto: Equipe SER

Prof. Ariel e prof. Felipe em suas respectivas oficinas para iniciantes na linguagem R, uma no campus do Valonguinho e outra no campus do Gragoatá.

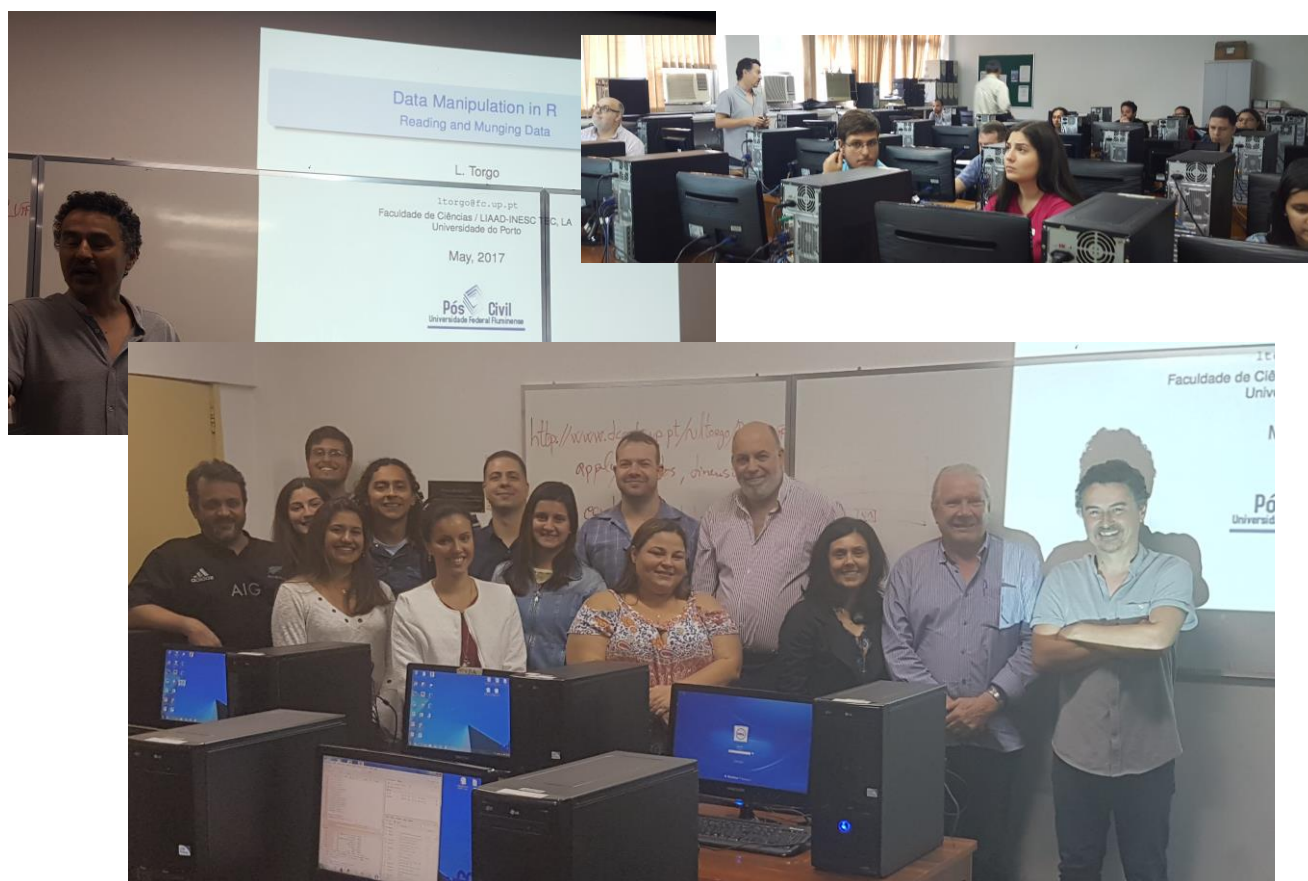


Sessão Pôster e Sessão Oral

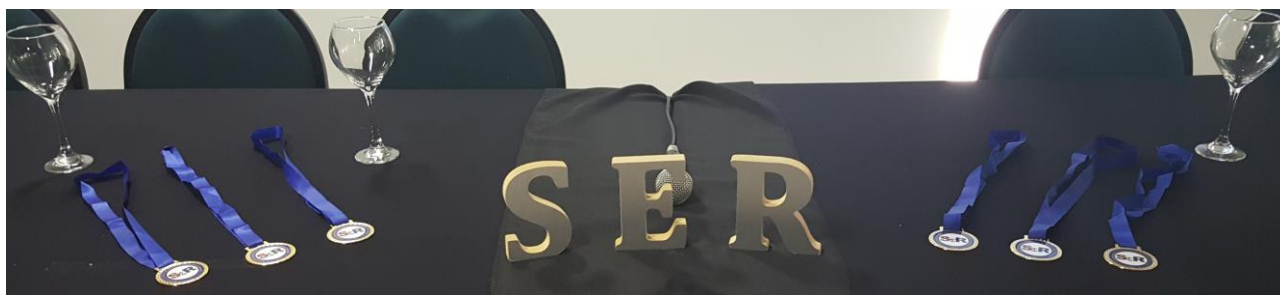


Atividades Complementares ao evento...

Minicurso de Data Mining with R com prof. Luís Torgo



Premiação



Tanto a sessão oral como a sessão pôster contaram com a premiação dos três melhores trabalhos. Nesta edição participaram 27 pôsteres e 21 artigos apresentados na sessão oral. De acordo com a avaliação da Comissão Científica do II SER, o melhor pôster é o trabalho intitulado *Impacto da Redução da quantidade de alternativas de um item do ENEM na estimação da proficiência do participante*, de autoria de Alexandre Jaloto e Natália Caixeta Barroso e como melhor artigo o trabalho intitulado *Esquema Operacional de baixo custo para verificação estatística de modelos numéricos de previsão do tempo*, de autoria de Nilza Barros da Silva e Natália Santos Lopes.



O SER contou com parcerias externas à UFF, da Escola Nacional de Ciências Estatísticas (ENCE), do Instituto de Matemática Pura e Aplicada (IMPA), da Sociedade Brasileira de Matemática Aplicada e Computacional (SBMAC), o apoio da FAPERJ através do seu Programa de Apoio a Eventos e do Ministério da Indústria Comércio Exterior e Serviços.

Participação de palestrante do Ministério da Indústria, Comércio Exterior e Serviços

Saulo Guerra e Paulo Felipe, representantes do MDIC, tiveram importante participação no II SER, mostrando ao público como a equipe de analistas do MDIC utiliza o R para resolver as demandas de análise de dados, produzindo inclusive um pacote para visualização de dados, o D3PlusR.



Encerrando o evento...



“O II SER veio com inovações sem perder o alto nível e qualidade, foram inseridas apresentações orais de trabalhos com premiação e minicursos, além das sessões pôster com premiação, palestras, teds e apresentação de blogs. Estamos felizes pelo realizado e esperamos vocês no III SER.” – prof. Orlando Longo – coordenador do Programa de Pós-Graduação em Engenharia Civil da UFF>

Nada disso seria possível sem a equipe de voluntários...

Nossos sinceros agradecimentos a todos os participantes, palestrantes, apoiadores, voluntários, professores das comissões organizadora e científica.



Coordenadora Geral do SER – Profa. Luciane F. Alcoforado

“Nós estudamos o passado, vemos todas as estruturas do passado que há e predizemos o futuro. Não usamos bola de cristal, usamos matemática” – Prof. Manuel Febrero – II SER 2017

“Big Data não diz respeito apenas ao volume de dados mas também a sua diversidade em fontes de diferentes naturezas.”
– Prof. Luís Torgo – II SER 2017



Redação: Profa. Luciane F. Alcoforado

Comissão Organizadora

Docentes

Luciane Ferreira Alcoforado - UFF – Presidente

Ariel Levy - UFF Vice-Presidente

Orlando Celso Longo - UFF

José Rodrigo de Moraes - UFF

Alex Laier Bordignon – UFF

Fabiano dos Santos Souza - UFF

Discentes

Raquel Rocha dos Santos

Noelle Santos Camello

Camila de Santana Delgado Simões

Leonardo Filgueira

Flávia Lamin

Comissão Científica

Manuel Febrero Bande - USC/ES

Wenceslao Gonzalez Manteiga - USC/ES

Luciane Ferreira Alcoforado - UFF

Orlando Celso Longo - UFF

Ariel Levy - UFF

Emil de Souza Sanchez Filho - UFF

Carlos Alberto Pereira Soares - UFF

Assed Naked Haddad - UFRJ

Maysa Sacramento de Magalhães - ENCE/IBGE

José Rodrigo de Moraes - UFF

Steven Dutt Ross - UNIRIO

Djalma Galvão Carneiro Pessoa - ENCE/IBGE

Pedro Costa Ferreira - FGV/IBRE

Jorge Passamani Zubelli – IMPA

Avaliadores Fase 1 e 2

Manuel Febrero Bande - USC/ES

Luciane Ferreira Alcoforado - UFF

Orlando Celso Longo - UFF

Ariel Levy - UFF

Maysa Sacramento de Magalhães - ENCE/IBGE

José Rodrigo de Moraes - UFF

Steven Dutt Ross - UNIRIO

Djalma Galvão Carneiro Pessoa - ENCE/IBGE

Jorge Passamani Zubelli – IMPA

Alexandre Souza e Silva – UNIRIO

Márcio Cataldi – UFF

Fabiano dos Santos Souza – UFF

Eduardo Camilo da Silva - UFF

Equipe SER

Elizete Gomes - Secretária

Camila de Santana Delgado Simões -

Leonardo Filgueira

Raquel Rocha dos Santos

Equipe LEVE

Noelle Santos Camello – coordenadora do cerimonial

Breno Quintanilha -cerimonial

Equipe de apoio ao evento

Alunos do curso de Turismo e Hotelaria da UFF

Thaís Brites Dias Valença

Mayahra Asheley Medeiros Rezende

Manuela Sales Delomongo

Thaís Guedes Silva

Vinicius de Almeida Coimbra

Pâmela Natividade dos Santos

Ricardo Macario

Marina de Mendonça Ferreira

Daniel Wendel Diniz Manoel

Ana Luiza Felgueiras de Souza Almeida

Mariana de Queiroz Guimarães

Gabriela Menezes Quitã Nogueira

Anna Luisa Ribeiro Grilo Almendro

Larissa Santos Benevides Carqueja

Luisa Almeida de Mello

Gabrielle Duarte Rego

Beatriz Grazielle Thomaz Alves

Claíssa Fonseca Rosa

Dandara Moreira Oliveira

Alunos do curso de Administração

Thamires Nascimento Veiga – planilhas de contato de autoridades

Alunos do curso de Estatística

Camila Simões – scripts do R para iniciantes

Leonardo Filgueira – scripts do R para iniciantes

Apoio IME

Alci Jorge – apoio técnico laboratório da Matemática

Heber Jorge – apoio técnico laboratório da Estatística

Leonardo Rafael dos Santos – apoio administrativo dos blocos G e H

Apoio Escola de Engenharia

Laboratório de Estruturas (sala de apoio, computadores, internet)

Núcleo de Mídia: Caio Melo da Silva e Matheus Gomes Bibiano

Apoio PPGEC

Laboratório de Informática - Leonardo de Oliveira Ibanez

Apoio PPGAD

Laboratório de Informática

Apoio Faculdade de Turismo e Hotelaria

Núcleo de Projetos (sala de apoio, computadores, telefonia e internet)

Apoio UNITEVE

Entrevistas, cobertura jornalística

Apoio PROEX, CEAX, Datauff

Certificados, inscrições, divulgação, programação visual

Apoio NAB

Carlos José de Oliveira e Silva – apoio técnico e operacional
Auditório e salas multiuso

Apoio Instituto de Computação

Hall do auditório para sessão pôster

Apoio IMPA

Convidados Internacionais da Espanha

Apoio ENCE

Coffee-breaks, exposição de livros

Parceria SBB net

Medalhas

Coquetel

Pimenta Góde – Francisca Paula

Apoio MDIC

Envio de palestrantes ao evento

Elaborado e Editado por Profa. Luciane F. Alcoforado

Revisão: Profs. Ariel Levy, Orlando C. Longo, José R. de Moraes

Participantes 194

Palestrantes 25

Autoridades 6

Professores 60

Equipe de apoio 35

Trabalhos premiados

Sessão de Comunicação Oral – categoria melhor artigo

1º. Colocado: Esquema Operacional de Baixo Custo para Verificação Estatística de Modelos Numéricos de Previsão do Tempo, de autoria de Nilza Barros da Silva e Natália Santos Lopes

2º. Colocado: Aplicação da Composição Probabilística de Preferências e do Índice de Gini à escolha de jogadores da Liga Inglesa de Futebol, de autoria de Luiz Octávio Gavião, Vitor Ayres Príncipe, Gilson Brito Alves Lima, Annibal Parracho Sant'Anna

3º. Colocado: Risco sistêmico na rede bancária brasileira: uma abordagem com Vine-cópula, de autoria de Andrea Ugolini e Miguel A. Rivera-Castro

Sessão Pôster – categoria melhor pôster

1º. Colocado: Impacto da Redução da Quantidade de Alternativas de um item do Enem na Estimação da Proficiência do Participante, de autoria de Alexandre Jaloto e Natália Caixeta Barroso

2º. Colocado: Shiny em Gráficos de Controle Estatístico de Processo, de autoria de Andréa Cristina Konrath, Rodrigo Gabriel de Miranda, Elisa Henning e Olga Maria Formigoni Carvalho Walter

3º. Colocado: The Drivers of Break-Even Inflation in Brazil: A Lasso Approach, de autoria de Daniel Karp, Luciano Vereda e Renato Lerípio.