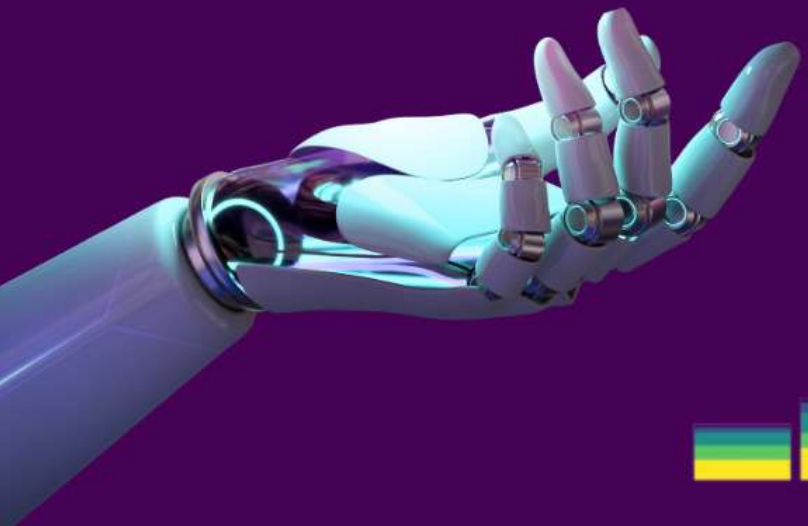
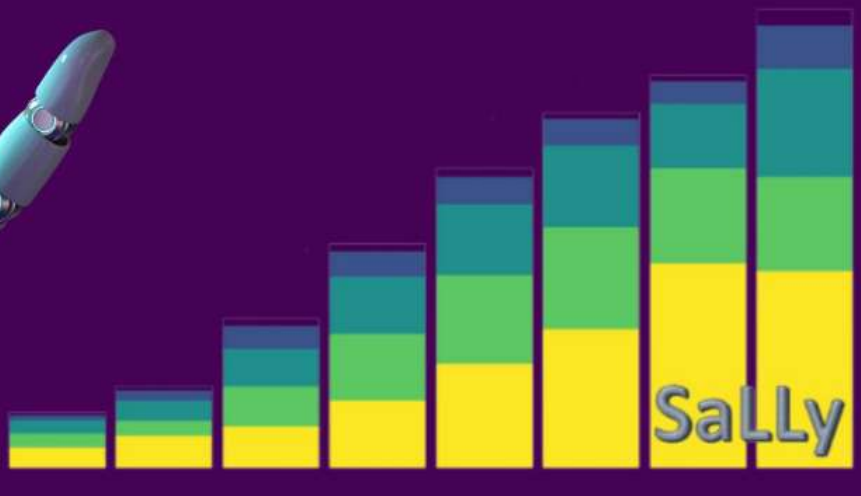
2nd SaLLy Day:

O PODER DA INTELIGÊNCIA ARTIFICIAL

Book of abstracts

30 de novembro de 2024





Livro de resumos do 2nd SaLLy Day

30 de novembro de 2024

Evento Online

Editado por
Paulo Canas Rodrigues
Beatriz Lopes
Maria Andreina Moreira
Universidade Federal da Bahia, Salvador, Brasil
Jonatha Sousa Pimentel
Universidade Federal de Pernambuco, Recife, Brasil

Web Design
João Vitor Rocha da Silva
SaLLy, UFBA, Salvador, Brasil

Design da Capa
Ana Caroline Pinheiro
SaLLy, UFBA, Salvador, Brasil

Citar como:
Rodrigues, P.C.; Lopes, B.; Moreira, M. A..; Pimentel, J.S.; Silva, J.V.R. ; Pinheiro, A. C. Livro de resumos do 2nd SaLLy Day, 2024.

Contents

Part I

Introdução

Bem Vindo ao 2st SaLLy Day!

Em nome da Comissão Organizadora e da Comissão Científica, é com grande entusiasmo que damos as boas-vindas a todos os amantes e entusiastas da estatística, da aprendizagem de máquina e da inteligência artificial a segunda edição do SaLLy Day, um evento emocionante e enriquecedor organizado pelo Statistical Learning Laboratory (SaLLy) da Universidade Federal da Bahia.

Neste dia de imersão intelectual, convidamos você a se juntar a nós para uma jornada repleta de descobertas, insights e aprendizados profundos no mundo da estatística e da inteligência artificial. O 2st SaLLy Day não é apenas um evento, mas uma oportunidade única de se conectar com especialistas renomados, colegas entusiasmados e mentes criativas que estão moldando o cenário da aprendizagem estatística.

A organização deste encontro foi realizada pelo SaLLy - Statistical Learning Laboratóry e o seu objetivo é reunir investigadores e profissionais, da academia e da indústria, que desenvolvam e apliquem métodos estatísticos e computacionais a inteligência artificial. Este evento proporcionará um fórum para compartilhar e discutir formas de melhorar o acesso ao conhecimento e promover colaborações interdisciplinares.

O programa científico inclui uma palestra de abertura, uma mesa redonda sobre o Papel da Estatística na Inteligência Artificial, um minicurso e uma sessão de pôster. Os expositores de pôster presentes no 2nd SaLLy Day tiveram a possibilidade de concorrer ao prêmio de melhor pôster.

Os organizadores gostariam de agradecer aos colaboradores que forneceram apoio para tornar esta organização possível. Agradecemos aos palestrantes, aos debatedores da mesa redonda, aos expositores de pôsteres, e a todos os participantes por sua contribuição para a confecção de um grande programa científico.

Obrigado a todos pela vossa contribuição!

Em nome do Comité do Programa Científico e do Comité Organizador Local,

Paulo Canas Rodrigues
Coordenador da Comissão Científica do 2st SaLLy Day

João Vitor R. Silva
Coordenador da Comissão Organizadora Local do 2st SaLLy Day

Comissão Organizadora

- João Vitor Rocha Silva (Coordenador), SaLLy, UFBA
- Ana Caroline Pinheiro da Cruz, SaLLy, UFBA
- Ana Carolina Andrade, SaLLy, UFBA
- Augusto Perin, SaLLy, UFBA
- Beatriz Lopes, SaLLy, UFBA
- Carlos Senra, SaLLy, UFBA
- Jonatha Sousa Pimentel, SaLLy, UFPE
- Kim Leone, SaLLy, UFBA
- Maria Andreina Moreira, SaLLy, UFBA
- Paulo Canas Rodrigues, SaLLy, UFBA

Comissão Científica

- Paulo Canas Rodrigues (Coordenador), SaLLy, UFBA
- Crysttian Paixão, SaLLy, UFBA
- João Vitor Rocha Silva, SaLLy, UFBA
- Olawale Awe, SaLLy, UFBA
- Rodrigo Bulhões, SaLLy, UFBA
- Valdério Anselmo Reisen, SaLLy, UFBA
- Vanda Lourenço, SaLLy, UFBA
- Vanessa Barros de Oliveira, SaLLy, UFBA

Part II

Programação Científica

Programação Científica

09h15-09h30: Abertura

09h30-10h30: Palestra de Abertura: Aprendizagem de máquina e as fronteiras com a Estatística e a Ciência de Dados, Prof. Dr. Luciano Rebouças Oliveira - Instituto de Computação, UFBA

10h30-10h45: Intervalo

10h45-12h15: Mesa Redonda - O Papel da Estatística na Inteligência Artificial. Jorge Mendes (Universidade Nova de Lisboa), Júnia Ortiz (Líder Técnica em Big Data e IA no Senai Cimatec), Téo Calvo (Instituto Aaron Swartz) e Wagner Bonat (UFPR). Moderador: João Vitor Rocha (SaLLy)

12h15-14h00: Intervalo para Almoço

14h00-17h15: Minicurso

- Transforme seus PDFs em Chatbots: Um Guia Completo para RAG com R e Python, Magno T. F. Severino (Insper)

17h30-18h30: Sessão de Postêres

18h30-18h45: Encerramento e Premiação

Part III

Palestra de abertura

Aprendizagem de máquina e as fronteiras com a estatística e a ciência de dados

Luciano Rebouças Oliveira[1,2]

[1] Universidade Federal da Bahia, Brasil
[2] IVISION - Laboratório de Pesquisa em Visão Inteligente, UFBA, Brasil
Email: lrebouca@ufba.br

Abstract

A palestra explora o campo da aprendizagem de máquina, destacando sua interseção com a estatística e a ciência de dados. Serão abordados conceitos fundamentais desses campos, como modelagem estatística e algoritmos de aprendizado, e como suas fronteiras estão se tornando cada vez mais tênues no contexto moderno. O foco será na importância da estatística para garantir a robustez e a interpretabilidade dos modelos de aprendizado, bem como na contribuição da ciência de dados para a aplicação prática em grande escala. Exemplos práticos e desafios atuais serão discutidos, tentando apresentar uma visão abrangente para aqueles que desejam entender como esses domínios se complementam.

Part IV

Minicurso

Transforme seus PDFs em chatbots: Um guia completo para RAG com R e python

Magno T. F. Severino[1]

Instituto de Ensino e Pesquisa, INSPER, São Paulo
E-mail: magnotairone@gmail.com

Resumo: Neste minicurso, você aprenderá a construir chatbots inteligentes capazes de responder a suas perguntas com base no conteúdo de documentos em formato PDF usando R e Python. Utilizando as mais recentes tecnologias de inteligência artificial, como grandes modelos de linguagem e RAG (Retrieval-Augmented Generation), você dará vida aos seus documentos, tornando-os acessíveis e interativos. Abordaremos desde os conceitos básicos de inteligência artificial generativa até a implementação prática de sistemas de RAG. Você aprenderá a preparar seus dados, criar representações numéricas (embeddings) dos textos, armazenar essas representações em bancos de dados vetoriais e, por fim, construir chatbots capazes de gerar respostas coerentes e informativas.

Part V

Mesa Redonda

O papel da estatística na inteligência artificial

Júnia Ortiz[1], Jorge Mendes[2], Wagner Bonat[3], Téo Calvo[4] e João Vitor da Silva[5]

[1] SENAI CIMATEC, Brasil
[2] Universidade Nova Lisboa, Portugal
[3] Universidade Federal do Paraná, Brasil
[4] Instituto Aaron Swartz, Brasil
[5] SaLLy - Statistical Learning Laboratory, UFBA, Brasil

Resumo: Nesta mesa redonda, especialistas discutirão como os conceitos estatísticos são fundamentais para o desenvolvimento e a validação de modelos de Inteligência Artificial. Serão abordados temas como modelagem de incertezas, estimativas probabilísticas, otimização de algoritmos e técnicas estatísticas aplicadas ao aprendizado de máquina, destacando como esses métodos garantem a robustez e a generalização dos modelos. Além disso, o debate explorará o papel da estatística na construção de modelos explicáveis e na avaliação da transparência e precisão dos resultados. Serão discutidas estratégias para lidar com viés e desvios em conjuntos de dados, promovendo uma IA mais justa e ética. A mesa será uma oportunidade para entender como a estatística sustenta o desenvolvimento de soluções de IA confiáveis e responsáveis.

Participantes:

- Júnia Ortiz (SENAI CIMATEC, Brasil)
  E-mail: junia.ortiz@gmail.com
- Jorge Mendes (Universidade Nova Lisboa, Portugal)
  E-mail: jorge.mendes@nms.unl.pt
- Wagner Bonat (Universidade Federal do Paraná, Brasil)
  E-mail: wbonat@ufpr.br
- Téo Calvo (Intituto Aaron Swartz, Brasil)
  E-mail: teo@teomewhy.org
- João Vitor da Silva (SaLLy, UFBA, Brasil)
  E-mail: rochajoaovitor@yahoo.com

Part VI

Sessão de Pôsteres

SP1: Modelos para dados de contagem com superdispersão: Uma aplicação em dados da malária

Abrantes Mussafo[1], José Govone[2] e Liciana Arruda[3]

[1] Universidade Estadual Paulista Júlio de Mesquita Filho, UNESP, Botucatu, São Paulo
Email: a.mussafo@unesp.br
[2] Universidade Estadual Paulista Júlio de Mesquita Filho, UNESP, Rio Claro, São Paulo
Email: js.govone@unesp.br
[3] Universidade Estadual Paulista Júlio de Mesquita Filho, UNESP, Botucatu, São Paulo
Email: liciana.silveira@unesp.br

Abstract

O estudo tem como objetivo comparar os modelos de Poisson e Binomial Negativo para a análise da taxa de incidência da malária. Os dados são de rotina, mensais, e foram coletados na província de Tete, Moçambique, em 15 distritos e 155 unidades sanitárias, no período de 2016 a 2022. A comparação foi realizada por meio do ajuste de modelos lineares generalizados, utilizando o específico de informação de Akaike (AIC), o parâmetro de dispersão, razão entre o desvio residual e o grau de liberdade e o diagnóstico do ajuste. Em relação aos resultados, o modelo de Poisson apresentou superdispersão igual a $\phi = 8.175219$ e uma razão entre a deviance residual e o grau de liberdade de 7.0603, ambos superiores a 1. Além disso, o AIC foi de 14060. Por outro lado, o modelo O Binomial Negativo teve um ajuste pressionando, com parâmetro de dispersão de $\phi = 1.15001$, uma razão entre o deviance residual e o grau de liberdade de 1.0539, AIC de $8766,9$ e o diagnóstico adequado. Conclui-se que o modelo Binomial Negativo é o mais adequado em situações de superdispersão nos dados. Além disso, pode-se afirmar que a taxa de analfabetismo e a temperatura mínima média são preditores da taxa de incidência da malária.

SP2: Modeling and forecasting intentional violent deaths (MVIs): Impact of actions in the years 2015 to 2023 in the state of Piauí

Dario Galvão[1] e Rita de Cássia de Lima Idalino[2]

[1] SEDUC, Piauí **Email**: dario1persa@gmail.com
[2] Universidade Federal do Piauí, UFPI, Teresina, Piauí
Email: rita@ufpi.edu.br

Abstract

This study investigates Intentional Violent Deaths in Piauí from 2015 to 2023. It analyzes historical data and employs predictive statistical models to understand trends in IVDs, noting significant fluctuations influenced by events like the COVID-19 pandemic and the rise of criminal organizations. The research highlights seasonal variations and spatial distributions, particularly in the urban peripheries of Teresina, identifying areas of high violence. The analysis shows variable trends in IVDs, especially in 2019, 2020, 2022, and 2023, with increased median values likely linked to economic crises and local conflicts. Monthly data indicates that cyclical factors, such as year-end festivities, also impact IVD incidence. A spatial analysis reveals regions in Teresina with a high potential for violence, underscoring the need for targeted interventions. A significant finding is the reduction of approximately 15.25% in IVDs from 2022 to 2023, reflecting the positive impact of SSP-PI's initiatives. This highlights the importance of data-driven security policies and strategic planning in public safety, enabling more effective resource allocation. The study advocates for an integrated approach to combat IVDs, combining quantitative and qualitative analyses, and emphasizes the need for ongoing monitoring and evaluation of public security.

SP3: Leak detection in water distribution networks through deep learning

Mariana G. M. Pereira[1], Luciano M. Queiroz[2], Karla P. Oliveira-Esquerre[3] e Enrique L. Droguett[4]

[1] Water and Sanitation Company, EMBASA, Bahia
Email: mariana.matias@embasa.ba.gov.br
[2] Department of Environmental Engineering, Graduate Program in Environment, Water and Sanitation, Polytechnic School of Engineering, Federal University of Bahia, Bahia
Email: lmqueiroz@ufba.br
[3] Department of Chemical Engineering, Graduate Program of Industrial Engineering, Polytechnic School of Engineering, Federal University of Bahia, Bahia
Email: karlaesquerre@ufba.br
[4] Department of Civil and Environmental Engineering Center for Reliability Science and Engineering, Garrick Institute for the Risk Sciences, University of California, USA
Email: eald@g.ucla.edu

Abstract

Water loss control is a priority for water supply companies around the world, given the scarcity of water resources. Leaks represent a large portion of the lost volume of water, and the precise and early detection of these anomalies in water supply networks remains a great challenge. Detection by inspection methods presents high monetary costs, limited response capacity, besides being a time-consuming and laborious activity. On the other hand, the detection approach using hydraulic models is difficult to implement, mainly in Brazil, due to the complex topology and the uncertainty in the hydraulic conditions of the water distribution networks. In this context, this research proposed the application of a Deep Learning algorithm to build a model for leak detection in water distribution networks. However, field hydraulic monitoring data presents high uncertainty, which makes it difficult to obtain accurate Machine Learning models. Therefore, a detailed exploratory and statistical data analysis was conducted to identify and preprocess the model's input data. The treatments performed, including the imputation of missing values through the training of secondary time series models and geoprocessing for locating the anomaly and labeling the dataset, made it possible to minimize the influence of: low data quality; uncertainty regarding the actual start of leaks; and operational regimes or maintenance on the network that generate changes in hydraulic parameters similar to leaks. One of the biggest challenges was the amount of data available, approximately 380 million records even after scope delimitation.

SP4: Adaptações do extreme gradient tree boosting (Xgboost) para base de dados desbalanceadas

Gabriel Almeida Ferreira[1] e Adriano Kamimura Suzuki[2]

[1] Universidade de São Paulo, USP, São Paulo
Email: gabrielalmeidaferreira@usp.br
[2] Universidade de São Paulo, USP, São Paulo
Email: suzuki@icmc.usp.br

Abstract

Nesse trabalho, foram estudadas maneiras de adaptar o XGBoost para bases de dados desbalanceadas. Isso foi feito de duas formas: por meio do balanceamento artificial dos dados e pela utilização de uma função de perda adequada. A respeito dos métodos de balanceamento dos dados, foram utilizados métodos de oversampling (Smote, ADASYN, Borderline Smote e Random Oversampling), undersampling (Edited Nearest Neighbor, Tomek Links e Random Undersampling) e abordagem mista (SmoteEEN). Ademais, uma nova função de perda foi proposta, utilizando como base a função de perda Weighted Focal Loss, que já foi utilizada como função de perda em redes neurais e, recentemente, no XGBoost. Essa modificação consiste em modelar as probabilidades estimadas por meio da ligação potência logito, o que adiciona flexibilidade à função de perda, permitindo assimetria para modelar as probabilidades. Portanto, a nova função de perda tem três hiperparâmetros: dois que vêm da função Weighted Focal Loss, para controlar os falsos negativos e para "direcionar o aprendizado aos exemplos difíceis de classificar", e um que veio da função de perda potência logito, responsável por introduzir assimetria na forma como as probabilidades são modeladas. Os modelos supramencionados foram testados em 13 conjuntos com diferentes graus de desbalanceamento, utilizando um procedimento de otimização de hiperparâmetros e um esquema de validação adequado para evitar vazamento dos dados. Em geral, observamos que os modelos estudados podem ser utilizados como alternativa ao XGBoost sem nenhuma modificação, já que houve aumento nas métricas de área sobre a curva precision recall e área sobre a curva ROC.

SP5: O uso de LightGBM para identificar autistas: uma revisão sistemática

Nayara Mota[1]

Universidade Católica, Salvador, Bahia
Email: nayara.mota@toryneuropsi.com.br

Abstract

Assim como a estatística convencional, a aprendizagem de máquina pode ser usada para analisar aspectos individuais e contextuais. O modelo de classificação LightGBM destaca-se por sua alta precisão e flexibilidade para manejar padrões não lineares. Assim, considera-se a ampliação de sua aplicabilidade para estudos preditivos de quadros clínicos com alta variabilidade, como o transtorno do espectro autista. Através de uma revisão sistemática nas principais bases de dados - IEEE, ACM, EMBASE e Google Acadêmico (os primeiros 10 registros), objetivou-se compreender o cenário científico quanto ao uso de LightGBM para identificar autismo. Realizou-se a seguinte busca por artigos: "lightGBM AND (autism or autistic)". No IEEE e no ACM, não houve resultados na categoria de artigos. Na MEDLINE, o único registro foi excluído por não referir-se ao autismo. No Google Acadêmico, após triagem e exclusão de registros que não fossem artigos ou que tivessem outros objetivos ou técnicas, foram incluídos 2 artigos à revisão. Entre estes, o uso de LightGBM, junto com a pesquisa aleatória para otimização de hiperparâmetros, promoveu alta precisão (>95) na classificação de autistas, em equivalência ao Autistic Spectrum Disorder Screening Test; assim como alta (> 99%) precisão, especificidade e sensibilidade na classificação de autistas, ao manejar dados psicofisiológicos do eletroencefalograma. Portanto, esta aplicabilidade da LightGBM na classificação de pessoas autistas ainda é incipiente. A substituição das técnicas estatísticas convencionais por aprendizagem de máquina, especificamente LightGBM, para a validação de técnicas diagnósticas do autismo - como questionários e neuroimagem - demandará o manejo de variáveis clínicas e sociais diversificadas, através da interdisciplinaridade.

SP6: Explainable artificial intelligence for defect detection in industrial manufacturing processes

Rodrigo Marcel Araujo Oliveira[1], Ângelo Márcio Oliveira Sant'Anna[2] e Paulo Henrique Ferreira[1]

[1] Polytechnic School, Federal University of Bahia, Bahia
Email: rodrigomarcel@ufba.br
[2] Polytechnic School, Federal University of Bahia, Bahia
Email: angelo.santanna@ufba.br
[3] Institute of Mathematics and Statistics, Federal University of Bahia, Bahia
Email: paulohenri@ufba.br

Abstract

In Industry 4.0, machine learning algorithms for non-linear pattern recognition are transforming defect detection in manufacturing, enabling enhanced quality monitoring, operational efficiency, and competitiveness. Explainable Artificial Intelligence (XAI) can facilitate understanding how the models make decisions and assist in tracking anomaly points. This research proposes an XAI framework for machine learning models in defect detection. This work applied the multinational industry's approach to the tire manufacturing process. Models such as random forest, gradient boosting decision tree, light gradient boosting machine, logistic regression, support vector machine, and multilayer perceptron were considered to evaluate the performance of tires in compliance with production standards. The selection of model parameters was based on the optimization technique using genetic algorithms. The random forest and logistic regression models achieved the best performances with an accuracy of 92% and 96%, respectively. The Local Interpretable Model-Agnostic Explanations (LIME) and SHapley Additive exPlanations (SHAP) methods were used to identify the relevant process variables. The approach presents the global and local interpretations of the results, allowing for a deeper understanding of how process variables discriminate the fault. This approach provides a relevant tool for quality control management in tire manufacturing industries.

SP7: Graph neural network-based anomaly detection in financial transactions associated with money laundering

Rodrigo Marcel Araujo Oliveira[1], Ângelo Márcio Oliveira Sant'Anna[2] e Paulo Henrique Ferreira[1]

[1] Polytechnic School, Federal University of Bahia, Bahia
Email: rodrigomarcel@ufba.br
[2] Polytechnic School, Federal University of Bahia, Bahia
Email: angelo.santanna@ufba.br
[3] Institute of Mathematics and Statistics, Federal University of Bahia, Bahia
Email: paulohenri@ufba.br

Abstract

The money laundering is a global threat, compromising the integrity of the financial system and risking the stability of the global economy. This work proposes the use of complex network techniques and presents a methodology to detect anomalies in financial transactions of individuals under investigation for suspected money laundering. The methodology involves creating various financial indicators, such as the average, sum of transaction values, and the number of transactions sent and received by each bank account. In this context, the account number represents the node in the directed graph. The Unifying Local Outlier Detection Methods via Graph Neural Networks (LUNAR) algorithm was used to recognize patterns in financial transactions and identify anomalies. The results highlight the model's effectiveness, with Silhouette score and Davies-Bouldin Index metrics of 1.59 and 0.83 achieved on the test set, respectively. This indicates that groups of anomalous and normal accounts are well represented in terms of similarity and dissimilarity. The results are promising and may assist in investigations by helping to identify potential groups of individuals involved in illicit activities, such as drug and arms trafficking, fraud, and scams.

SP8: O Método bootstrapping em regressão logística para previsão de diabetes

Carla Patrícia de Carvalho Oliveira[1], Álbaro Ramon Paiva Sanz[2], Ursilandia de Carvalho Oliveira[3], Liciana de Arruda Silveira[4] e Viriato Campelo[5]

[1] Universidade Estadual Paulista Júlio de Mesquita Filho, UNESP, Botucatu, São Paulo
Email: carla.patricia@unesp.br
[2] Universidade Estadual da Paraíba, UEPB, Campina Grande, Paraíba
Email: albaropaiva@gmail.com
[3] Instituto Federal de Educação, Ciência e Tecnologia do Maranhão, IFM, Codó, Maranhão
Email: ursilandia.oliveira@ifma.edu.br
[4] Universidade Estadual Paulista Júlio de Mesquita Filho, UNESP, Botucatu, São Paulo
Email: liciana.silveira@unesp.br
[5] Universidade Federal do Piauí, UFPI, Piauí
Email: viriato.campelo@bol.com.br

Abstract

Existem inúmeras aplicações que utilizam regressão logística em amostras obtidas por reamostragem, aproveitando a capacidade dos métodos de lidar favoravelmente com a alta demanda computacional. Neste contexto, aplicamos a técnica de bootstrap para prever o início do diabetes com base em um conjunto de dados de saúde. O conjunto de dados é composto por 768 registros de pacientes do sexo feminino, cada um caracterizado por oito atributos de saúde e uma variável binária indicando a presença (1) ou ausência (0) de diabetes. Ressalta-se que ajustou-se um modelo de regressão logística através de um modelo linear generalizado para dados binários, com o intuito de estimar a probabilidade de ocorrência de diabetes em função da glicose. Desse modo, apresentamos através dos intervalos de confiança de 95% dos parâmetros desconhecidos do modelo do modelo logístico e odds ratio, para testar a hipótese de que a prevalência de diabetes não depende da glicose. Assim, utilizamos métodos clássicos e de bootstrap, incluindo abordagens paramétricas e não-paramétricas. Observamos que ao comparar os métodos bootstrap e a regressão logística clássica os resultados foram bastante similares. Conforme observado, o método bootstrap é uma técnica muito eficiente para avaliar a variabilidade dos coeficientes do modelo e apresentar estimativas robustas, prioritariamente em amostras pequenas ou em distribuições de variáveis desconhecidas.

SP9: Análise espaço-temporal da atividade de raios no Brasil de 2020 até 2022

Beatriz Lopes[1] e Paulo Canas Rodrigues[2]

[1] SaLLy, UFBA, Bahia
Email: beatrizlopes@ufba.br
[2] SaLLy, UFBA, Bahia
Email: paulocanas@gmail.com

Abstract

O Brasil é considerado o país com maior incidência de raios no mundo. De acordo com estudos realizados pelo ELAT (Grupo de Eletricidade Atmosférica), caem em média 77,8 milhões de descargas no Brasil anualmente. A localização na região tropical, combinada com altas temperaturas e elevada umidade do ar, cria condições ideais para a formação desse fenômeno de consequências proporcionais a sua intensidade. Este trabalho tem como objetivo analisar a distribuição espaço-temporal da atividade de raios no Brasil, utilizando técnicas de Análise Exploratória de Dados Espaciais (AEDE) para interpretar os dados. Os materiais utilizados neste estudo foram obtidos do acervo virtual da Universidade Federal de Itajubá (UNIFEI), que disponibiliza registros de raios para toda a América Latina. Essas informações são coletadas pelo aparelho GOES-16 abordo do sensor Geostationary Lightning Model (GLM), operado pela NASA. O período de estudo (2020-2022) corresponde ao intervalo disponível para download no acervo. A fim de compreender a distribuição espacial dos raios no território brasileiro, foram utilizados os métodos estatísticos I de Moran Global e I de Moran Local. Através do I de Moran Global, foi possível identificar a existência de uma autocorrelação espacial do fenômeno. Já o I de Moran Local (LISA) confirmou a existência de uma autocorrelação positiva, sendo evidenciada visualmente em gráfico de dispersão e cartograma. Esses resultados indicam que municípios com alta concentração de raios tendem a estar próximos geograficamente uns dos outros, além disso, representam a maior parte da região estudada.

SP10: Aprimoramento da detecção de câncer em bases de imagens limitadas: Uma abordagem baseada em ensemble e aumento de dados

Fernando Moraes[1], Adriano Suzuki[2], Francisco Louzada Neto[3] e Ricardo Rocha[4]

[1] Universidade de São Paulo, USP, São Paulo
Email: fernandohumberto2009@hotmail.com
[2] Universidade de São Paulo, USP, São Paulo
Email: suzuki@icmc.usp.br
[3] Universidade de São Paulo, USP, São Paulo
Email: louzada@icmc.usp.br
[4] Universidade Federal da Bahia, UFBA, Bahia
Email: ricardo8610@gmail.com

Abstract

Modelos de Deep Learning são promissores na análise de imagens, mas exigem muitos dados, dificultando seu uso em bases médicas limitadas. Em dados pequenos e desbalanceados, técnicas como transferência de aprendizado, ensemble, aumento de dados e reamostragem (undersampling e oversampling) melhoram a classificação de imagens. Este estudo aplica uma abordagem de ensemble com diferentes pesos para aprimorar a predição de classes minoritárias em dados oncológicos de câncer de pele e mama, obtendo melhores métricas com reamostragem e aumento de dados.

SP11: Modelo casual decision tree: Uma aplicação na avaliação do impacto heterogênico da campanha de vacinação da dengue

Diego Santos Souza[1] e Paulo Canas Rodrigues[2]

[1] Universidade Federal da Bahia, UFBA, Bahia
Email: seg.diego1355@gmail.com
[2] Universidade Federal da Bahia, UFBA, Bahia
Email: paulocanas@gmail.com

Abstract

As vacinas são um método de controle de epidemias, onde por meio de estudos clínicos randomizados controlados, são avaliadas a sua eficácia em um nível individual, porém os estudos clínicos citados não conseguem avaliar um programa de vacinação a nível populacional, e métodos quasi-experimentais precisam ser usados. Neste trabalho analisamos a casualidade do impacto heterogêneo da campanha de vacinação da dengue na redução de sua incidência utilizando o modelo Casual Decision Tree.

SP12: Análise temporal e espacial da inadimplência das famílias brasileiras

Maria Moreira[1], Vanessa Barros[2] e Paulo Canas Rodrigues[3]

[1] SaLLy, UFBA, Bahia
Email: maria.andreina@ufba.br
[2] SaLLy, UFBA, Bahia
Email: vbarrosoliveira@gmail.com
[3] SaLLy, UFBA, Bahia
Email: paulo.canas@ufba.br

Abstract

Em um país marcado por flutuações econômicas e desigualdades de renda, o Brasil encontra-se entre os países com maior taxa de inadimplência familiar global. Esse cenário evidencia tanto questões individuais quanto estruturais, como restrições no acesso a recursos básicos de sobrevivência. Este trabalho se propõe a contribuir com análises sobre as famílias em inadimplência nas capitais dos estados brasileiros e busca trazer previsões sobre esse comportamento no país. Para isso, foram empregadas técnicas gráficas de mapeamento, análise de clusters e também métodos comparativos dos modelos de ajuste e previsão de séries temporais, Modelo Autorregressivo Integrado de Médias Móveis(ARIMA), Exponential Smoothing(ETS) e Prophet, com dados da Confederação Nacional do Comércio de Bens, Serviços e Turismo (CNC) de janeiro de 2010 a dezembro de 2023. Os resultados obtidos pretendem trazer informações relevantes sobre a inadimplência no Brasil, destacando tendências regionais e projeções futuras. A análise espaço-temporal desempenhou um papel crucial na compreensão desses padrões. Em seus resultados, foi possível compreender a dinâmica da inadimplência em cada localidade; por exemplo, foi possível identificar que tanto a região Nordeste quanto a Sudeste tiveram maiores taxas e, em alguns períodos, também apresentaram um certo padrão. Já a análise de séries temporais trouxe importantes informações sobre previsões de inadimplência no país; os resultados indicaram que, de forma geral, modelos ARIMA e ETS foram melhores para calcular as previsões. Esses resultados podem auxiliar no direcionamento de políticas públicas estratégias financeiras que visem abordar com precisão os desafios enfrentados pelas famílias brasileiras no contexto do endividamento.

SP13: Comparison between computer methods in the automatic detection of Alzheimer's disease

Andriana Campanharo[1], Mário Vicchietti[2], Luiz Betting[3] e Fernando Ramos[4]

[1] Universidade Estadual Paulista Júlio de Mesquita Filho, UNESP, Botucatu, São Paulo
Email: andriana.campanharo@unesp.br
[2] Universidade Estadual Paulista Júlio de Mesquita Filho, UNESP, Botucatu, São Paulo
Email: mario.lucas@unesp.br
[3] Universidade Estadual Paulista Júlio de Mesquita Filho, UNESP, Botucatu, São Paulo
Email: luiz.betting@unesp.br
[4] Instituto Nacional de Pesquisas Espaciais, INPE, São José dos Campos, São Paulo
Email: fernando.ramos@inpe.br

Abstract

Alzheimer's disease (AD) is a progressive disease in which nerve cells degenerate, leading to memory loss, learning difficulties, spatial and temporal disorientation, mood swings, personality changes and even loss of motor skills. Since there is no cure for Alzheimer's disease, its diagnosis is the best alternative for starting treatments that slow the progression of the disease. In this scenario, electroencephalography (EEG) has gained attention as it is a non-invasive technique that can measure the electrical potential emanating from neuronal activity. In recent decades, several groups of scientists have proposed to use computerized methods to analyze time series in EEG signals from patients with and without Alzheimer's disease to automatically identify the disease. The aim of this work is to compare the robustness of different computer methods for classifying patients with and without Alzheimer's disease using an EEG database.

SP14: Early detection of Alzheimer's disease using complex network analysis

Mário L. Vicchietti[1], Maria C. de Cola[2], Angelo Quartarone[3], Fernando M. Ramos[4] e Andriana S. L. O. Campanharo[5]

[1] Universidade Estadual Paulista Júlio de Mesquita Filho, UNESP, Botucatu, São Paulo
Email: mario.lucas@unesp.br
[2] IRCCS Centro Neurolesi, Messina, Italy
Email: mariacristina.decola@irccsme.it
[3] IRCCS Centro Neurolesi, Messina, Italy
Email: angelo.quartarone@irccsme.it
[4] Instituto Nacional de Pesquisas Espaciais, INPE, São José dos Campos, São Paulo
Email: fernando.ramos@inpe.br
[5] Universidade Estadual Paulista Júlio de Mesquita Filho, UNESP, Botucatu, São Paulo
Email: andriana.campanharo@unesp.br

Abstract

Alzheimer's disease (AD), the most prevalent form of dementia, remains incurable. The early detection in its initial stage, known as Mild Cognitive Impairment (MCI), is crucial for enhancing patients' quality of life. Electroencephalography (EEG), a non-invasive diagnostic tool, has been used in studying and identifying AD. This research explores the changes caused by AD at different stages by analyzing EEG sub-bands using a novel complex network approach, the Quantile Graph method. EEG signals were collected from 20 healthy controls (HC), 37 MCI patients, and 48 AD patients, and three topological measures were extracted as features for classifying and distinguishing the health conditions of the subjects. Multivariate Analysis of Variance revealed significant statistical differences among the three groups. Classification results using a Support Vector Machine achieved accuracies of 94%, 93%, and 90% for HC vs. AD, HC vs. MCI, and MCI vs. AD, respectively. In conclusion, the proposed method demonstrates significant potential for distinguishing these groups from a complex network perspective.

Index