

Ementas das Disciplinas

- **Obrigatórias**

1) Introdução a Ciência de dados

Fundamentos da Ciência de Dados combina três perspectivas: pensamento inferencial, pensamento computacional e relevância no mundo real. O curso ensina conceitos e habilidades críticas de análise descritiva e inferencial a partir de ferramentas computacionais.

Referências:

1. Golemund, G. (2014). Hands-On Programming with R. O'Reilly. <https://rstudioeducation.github.io/hopr/>;
2. Wickham, H.; C, etinkaya-Rundel, M.; Golemund, G. (2023). R for data science. O'Reilly; 2nd Edition. <https://r4ds.hadley.nz/>;
3. Peng, R. D. (2016). R programming for data science. Victoria, BC, Canada: Leanpub. <https://bookdown.org/rdpeng/rprogdatascience/>
4. Wickham, H.; Navarro, D.; Thomas Lin Pedersen, T. L. (2023). Ggplot2: Elegant Graphics for Data Analysis. Springer; 3rd Edition. <https://ggplot2-book.org/>;
5. Carson Sievert (2020). Interactive web-based data visualization with R, plotly, and shiny. CRC Press. <https://plotly-r.com/>

2) Probabilidade Aplicada à Ciência de Dados

Introdução: Probabilidade. Simulação de Monte Carlo. Probabilidade condicional, independência de eventos com exemplos no software R. Variáveis aleatórias. Variáveis aleatórias discretas e contínuas. Momentos via solução analítica e por simulação de Monte Carlo. Transformação de variáveis aleatórias. Aplicação dos modelos probabilísticos em diversas áreas do conhecimento.

Referências:

1. Ross, S. Probabilidade: Um curso moderno com aplicações. 8a ed. Bookman: 2010.
2. Meyer, P. L. Probabilidade: Aplicações e Estatística. 2 ed. Rio de Janeiro: Livros Técnicos e Científicos. Editora S.A., 1984.

3) Inferência Aplicada à Ciência de Dados

Introdução à inferência estatística computacional. Grau e certeza sobre o que está acontecendo: Estimativas e Intervalos. Interpretando o intervalo de confiança com uso da Simulação de Monte Carlo. Estimação Pontual. Otimizando a função de verossimilhança: uso do R. Testes de Hipóteses para uma População. Testes de Hipóteses para duas Populações. Análise de Variância. Alguns testes usuais paramétricos e não paramétricos. Aplicação dos procedimentos de inferência em diversas áreas do conhecimento. Uso do Bootstrap paramétrico e não paramétrico.

Referências:

1. Bolfarine, H., Sandoval, M.C. Introdução à Inferência Estatística. Coleção Matemática Aplicada – Sociedade Brasileira de Matemática, 2001.
2. Hogg, R., Craig, A. McKean, J. W., Introduction to mathematical statistics. 6 ed. Upper Saddle River. N. Jersey, 2005.
3. Casella G. e Berger, R. L. Inferência Estatística. São Paulo: Cengage Learning, 2014.

4) Introdução à Programação em Ciência de Dados

Ambientes de programação em Ciência de Dados. Práticas de programação em Ciência de Dados. Resolução de Problemas via Computador. Estruturas Básicas. Comandos de Seleção e de Repetição. Procedimentos e Funções. Arquivos. Estruturas de dados. Introdução a bibliotecas para Ciência de Dados.

Referências:

1. Charles Severance. Python for everybody. CreateSpace Independent Publishing Platform. 2016

5) Engenharia de Dados

Tipos de dados. Qualidade de dados. Pré-processamento. Medidas de similaridade e dissimilaridade. Formatos de arquivos e armazenamento de dados. Limpeza e preparação dos dados. Organização dos dados. Agregação e operação de agrupamento. Redução de dimensionalidade.

Referências:

1. Tan, P.-N., Steinbach, M., Karpatne, A., & Kumar, V. (2018). Introduction to Data Mining (2nd Edition) (2nd ed.). Pearson.
2. Zaki, M. J., & Jr., W. M. (2020). Data Mining and Machine Learning: Fundamental Concepts and Algorithms. Cambridge University Press.
<http://www.dataminingbook.info/>
3. McKinney, W. (2017). Python for Data Analysis: Data Wrangling with Pandas, NumPy, and IPython (2nd ed.). O'Reilly Media, Inc.

6) Modelos Preditivos Lineares

Diagrama de dispersão. Método dos mínimos quadrados. Igualdade entre os estimadores de máxima verossimilhança e método dos mínimos quadrados. Necessidade de hipóteses para a componente de erro do modelo. Análise de Resíduos. Identificação de valores influentes. Transformação de variáveis. Variáveis indicadoras. Multicolinearidade. Seleção do melhor modelo de regressão. Aplicação dos procedimentos de regressão Linear em diversas áreas do conhecimento.

Referências:

1. Julian J. Faraway - Linear Models with R, Chapman & Hall/CRC Texts in Statistical Science, 2004.
2. MONTGOMERY, D. C.; PECK, E. *Introduction to linear regression analysis*. 3rd Edition. New York: Wiley-Interscience, 2001.

7) Elementos de Aprendizado de Máquina

Introdução aos Modelos Lineares Generalizados: Família exponencial. Inferência. Regressão para dados de contagem (Poisson), modelos com OFFSET. Regressão para respostas binárias (Logística), análise de sensibilidade e especificidade, curva ROC. Conceitos de Superdispersão e Subdispersão. Splines (Regressão Polinomial Parcial), GAM (Modelos Aditivo Generalizados), CART (Árvores de Regressão e Classificação). Uso do Bootstrap paramétrico e não paramétrico.

Referências:

1. CORDEIRO, G. M., *Modelos Paramétricos*. 16º SINAPE , ABE – Associação Brasileira de Estatística.
2. COSTA, Marcelo A.; Tópicos em Ciência dos Dados – Introdução dos Modelos Paramétricos e suas aplicações utilizando o R. Editora Bonecker, 1ª Edição, 2020.
3. Demétrio, C. G. B., *Modelos Lineares Generalizados em Experimentação Agrônômica*, 2002.
4. Dobson, A. J., *An Introduction to Generalized Linear Models*, Chapman & Hall, 1990.
5. VENABLES, W. N., SMITH, D. M. *An Introduction to R*. technical manual.

8) Ética e Fatores Humanos em Ciência de Dados

Legitimidade na tomada de decisão. Noções Relativas de Justiça
Causalidade. Justiça e Discriminação. Transparência. Responsabilidade. Auditabilidade.
Legislação.

Referências:

1. M. Quinn, Ethics for the Information Age, 7th Ed., Pearson, 2016.

9) Metodologia de Pesquisa em Ciência de Dados

Métodos Científicos. As Etapas da Pesquisa em Ciência de Dados. Definição e elaboração da monografia. Exemplos de problemas aptos a constituírem uma monografia em Ciência de Dados.

- **Optativas da Área de Concentração em Estatística:**

1) Séries Temporais e Aprendizado Dinâmico

Análise exploratória de séries temporais. Regressão em séries temporais. Métodos de decomposição de séries temporais. Métodos de alisamento exponencial. Modelos ARIMA. Modelos de Espaço de Estados. Aplicação dos procedimentos de Séries Temporais em diversas áreas do conhecimento: estudo de casos.

Referências:

1. R. H. Shumway, D. S. Stoffer - Time Series Analysis and its Applications: with R examples, New York: Springer, 2006.
2. B. L. Bowerman, R. T. O'Connell - Forecasting and Time Series: an applied approach, 3rd. ed., Wadsworth, Inc, 1993.
3. W. A. Woodward, B. P. Sadler, S. Robertson - Time Series for Data Science: Analysis and Forecasting. Chapman & Hall/CRC Texts in Statistical Science, 2022.

2) Aprendizado de Máquina Não-Supervisionado com enfoque Estatístico

Técnicas de sintetização de dados multivariados: análise de componentes principais, análise fatorial, análise discriminante. Análise de variância multivariada. Análise de cluster: K-means, DBSCAN, Gaussian Mixture Models, Clusterização Hierárquica.

Referências:

1. Dillon, W. R. , Goldstein, M. Multivariate Analysis Methods and Application. New York: John Wiley, 1984.
2. Hair, JR., J. F., Anderson, R. E., Tatham, R. L., Black, W. Análise Multivariada de Dados. São Paulo: Editora Bookman, 2005.
3. Jobson, J. D. Applied Multivariate Data Analysis. vol I. e II, New York: Springer Verlag, 1996.
4. Johnson, R. A., Wichern, D. W. Applied Multivariate Statistical Analysis. New Jersey: Prentice Hall, Inc, 2002.
5. Mingoti, S. A. Análise de Dados Através de Métodos de Estatística Multivariada: Uma Abordagem Aplicada, Belo Horizonte: Editora UFMG, 2005.

3) Aprendizado de Máquina Supervisionado com enfoque Estatístico

Análise de Regressão do ponto de vista preditivo. Função risco, estimação da capacidade de generalização dos modelos. Seleção de Modelos. Regressão Lasso e Ridge. Métodos de vizinhos mais próximos e árvores de regressão. Modelos de classificação, logística, naive Bayes, SVM e métodos plug-in.

Referências:

1. James, G., Witten, D., Hastie, T. e Tibshirani, R. An Introduction to Statistical Learning, with Applications in R, Springer 2013. Disponível gratuitamente em <http://www-bcf.usc.edu/~gareth/ISL/>
2. Hastie, T., Tibshirani, R. e Friedman, J. The Elements of Statistical Learning. New York: Springer, 2009. Disponível gratuitamente em <http://statweb.stanford.edu/~tibs/ElemStatLearn/>
3. Bishop, C. M. Pattern Recognition and Machine Learning. New York: Springer, 2006.
4. Wasserman, L. All of Statistics. New York: Springer, 2004.
5. Wasserman, L. All of Nonparametric Statistics. New York: Springer, 2006.

4) Estatística Computacional e Otimização

Computação estatística: organização, armazenamento, manipulação de dados e arquivos; Geração de números aleatórios; Simulação; escrevendo funções no R. Programação Linear e não linear com uso do R e Solver/Excel. Previsão de Demanda e políticas de estoque. Aplicações no setor da indústria e mercado. Uso do Bootstrap paramétrico e não paramétrico.

Referências:

1. Paulo Cortez - Modern Optimization with R, 2014, Springer.

2. Sheldon Ross - Probabilidade, 2010, Bookman.
3. Sheldon Ross - Simulation, 2012, Academic Press.

5) Redes Neurais Artificiais (Aprendizado Profundo)

Deep Learning. Fundamentos de Deep Learning: Inteligência Artificial, Aprendizado de Máquina e Deep Learning. Princípios do Aprendizado de Máquina (Machine Learning). Fundamentos das Redes Neurais Artificiais. Aplicações de Redes Neurais Artificiais. Conceitos e aplicações de Aprendizado de Máquina. Deep Learning para visão computacional. Deep Learning para processamento de textos. Melhores práticas para aplicações avançadas de Deep-Learning.

Referências:

1. CHOLLET, François; ALLAIRE, Joseph J. Deep Learning with R, Manning Publications, 2018.
2. HAYKIN, Simon. Neural networks and learning machines, 3/E. Pearson Education India, 2009.
3. HASTIE, Trevor et al. The elements of statistical learning: data mining, inference, and prediction. New York: springer, 2009.
4. GOODFELLOW, Ian; BENGIO, Yoshua; COURVILLE, Aaron. Deep learning. MIT press, 2016.

- **Optativas da Área de Concentração em Computação:**

1) Visualização de Dados

Fundamentos de percepção visual. Relacionamentos quantitativos. Padrões visuais. Técnicas de interação analítica. Construção de painéis (dashboards). Teoria de cores. Avaliação de visualizações

Referências:

1. Spence, Robert. Information visualization. Vol. 1. New York: Addison-Wesley, 2001.
2. Cleveland, William S. The elements of graphing data. Wadsworth Publ. Co., 1985.
3. Few, Stephen. "Show me the numbers." Analytics Pres (2004).

2) Mineração de Dados

Descoberta de conhecimento em bases de dados. Análise exploratória de dados. Padrões frequentes. Correlações. Associações. Avaliação.

Referências:

1. M. J. Zaki and W. Meira Jr. Data Mining and Machine Learning: Fundamental

Concepts and Algorithms. Cambridge University Press, 2020.

3) Aprendizado de Máquina Supervisionado Computacional

Modelos Lineares para Regressão e Classificação. Regressão logística. Avaliação de modelos preditivos. Modelos baseados em árvores. Modelos aditivos (ensembles). Máquinas de Vetor Suporte. Truque do Kernel. Regularização e Espaços de Hilbert

Referências:

1. The Elements of Statistical Learning: Data Mining, Inference, and Prediction. Second Edition. Trevor Hastie. Robert Tibshirani. Jerome Friedman

4) Aprendizado de Máquina não Supervisionado Computacional

Agrupamento baseado em representantes. Agrupamento por densidade. Agrupamento espectral. Kernels em agrupamentos. Decomposição de matrizes. Avaliação de modelos.

Referências:

1. Data Clustering: Algorithms and Applications (CRC Press), Ed. Charu Aggarwal, Chandan Reddy, 2014

5) Redes Neurais e Aprendizado Profundo Computacional

Introdução às Redes Neurais. Programação Diferenciável. Teorema da Aproximação Universal. Redes Neurais para Regressão. Redes Neurais para Classificação. Arquiteturas de redes neurais. Funções de ativação. Otimização para Redes Neurais. Regularização para Redes Neurais

Referências:

1. Aston Zhang, Zachary C. Lipton, Mu Li, Alexander J. Smola. Dive into deep learning. Cambridge University Press, 2023.

6) Recuperação da Informação

Acesso a dados textuais. Representação de consultas. Representação de documentos. Normalização (codificação; tokenização). Análise (conteúdo informacional; equivalência léxica, fonética, e semântica; colocações; escopo). Indexação (estruturas de dados; algoritmo básico). Aprendizado de modelos

Referências:

1. Introduction to Information Retrieval, by C. Manning, P. Raghavan, and H. Schütze. Cambridge University Press, 2008.
2. Search Engines: Information Retrieval in Practice, by B. Croft, D. Metzler, and T. Strohman. Pearson, 2009.
3. Modern Information Retrieval, by R. Baeza-Yates and B. Ribeiro-Neto. Addison-

Wesley Professional, 2011.

4. Information Retrieval: Implementing and Evaluating Search Engines, by S. Büttcher, C. Clarke, and G. Cormack. MIT Press, 2016.
5. Text Data Management: A Practical Introduction to Information Retrieval and Text Mining, by C. Zhai and S. Massung. Morgan & Claypool, 2016

7) Sistemas de Recomendação

Caracterização. Recomendação colaborativa. Recomendação baseada em conteúdo. Recomendação baseada em conhecimento. Recomendação híbrida. Tópicos avançados.

Referências:

1. Recommender Systems: An Introduction, Dietmar Jannach, Markus Zanker, Alexander Felfernig, Gerhard Friedrich. Cambridge University Press, 2010.
2. Recommender Systems Handbook, Francesco Ricci, Lior Rokach, Bracha Shapira, Paul B. Kantor. Springer, 2015.
3. Recommender Systems: The Textbook, Charu C. Aggarwal. Springer, 2016.
4. Statistical Methods for Recommender Systems, Deepak K. Agarwal, Bee-Chung Chen. Cambridge University Press, 2016

8) Processamento de Linguagem Natural

Fundamentos de linguística e processamento de linguagem natural. Metodologia de modelagem paramétrica e não-paramétrica a partir de dados. Tipos, técnicas e algoritmos de aprendizado profundo .

Referências:

1. Alexander Clark, Chris Fox, Shalom Lappin. The Handbook of Computational Linguistics and Natural Language Processing. Wiley-Blackwell, 2013.

9) Aprendizado de Máquina para Imagem e Visão Computacional

Convolução, padding. Convolução strided, same, em imagens RGB, exemplo de uma camada, notação, exemplo de uma CNN, pooling. Estudos de caso: LeNet-5, AlexNet, VGG, Convoluções 1x1, Inception e ResNets. Comparação das arquiteturas. Fundamentos de Visão Computacional. Segmentação Semântica. Detecção. Detecção Yolo. PDI e Open Set Recognition. Vision Transformers

Referências:

1. Aston Zhang, Zachary C. Lipton, Mu Li, Alexander J. Smola. Dive into deep learning. Cambridge University Press, 2023.

10) Aprendizado de Máquina para Jogos Digitais

Introdução a jogos digitais, agentes inteligentes, busca em espaço de estados, busca adversarial, processos de decisão de Markov, aprendizado por reforço, aprendizado por reforço profundo.

Referências

1. Reinforcement Learning: An Introduction. Richard S. Sutton and Andrew G. Barto. 2nd Edition. MIT Press, Cambridge, MA, 2018
2. Artificial Intelligence – A Modern Approach. Stuart Russell and Peter Norvig, 4th Edition, Prentice Hall, 2020
3. An Introduction to MultiAgent Systems. Michael Wooldridge, 2nd Edition, Wiley Publishing, 2009.
4. Fundamentos de Visão Computacional. Segmentação Semântica. Detecção. Detecção Yolo. PDI e Open Set Recognition. Vision Transformers

11) Aprendizado de Máquina para Sinais e Áudio

Amostragem de Áudio. Espectrograma e Transformada de Fourier. Sincronização de Músicas. Segmentação Semântica. Identificação e Recuperação. Reconhecimento de Tempo. Decomposição de Matrizes de Áudio. Reconhecimento de Acordes.

Referências:

1. Meinard Muller. Fundamentals of Music Processing. Springer. 2021

12) Práticas de Engenharia de Software para Ciência de Dados

Processos de Software. Requisitos de Software. DevOps. Integração contínua. Entrega contínua e feature branches. MLOps.

Referências:

1. Marco Tulio Valente. Engenharia de Software Moderna: Princípios e Práticas para Desenvolvimento de Software com Produtividade, 2020.

13) Gerenciamento Operacional em Ciência de Dados

Implantação de Pipelines de Ciência de Dados. Coleta e Armazenamento de Dados. Monitoração de métricas (Treinamento Supervisionado). Monitoração de métricas (Treinamento Não Supervisionado). Drift de Modelos. Servindo Modelos.

Referências

1. Andriy Burkov. Machine Learning Engineering.
2. David Sweenor, Steven Hillion, Dan Rope, Dev Kannabiran, Thomas Hill, Michael O'Connell. MLOPS: Operationalizing Data Science.

14) Processamento de Dados Massivos

Noções de sistemas distribuídos. Escalabilidade, replicação, particionamento e transações. Modelos de dados e data lakes/warehouses. Processamento de dados em paralelo. Processamento em Batch vs Streaming. Estratégias de Treinamento distribuído

Referências:

1. Martin Klepmann. Designing Data-Intensive Applications: The Big Ideas Behind Reliable, Scalable, and Maintainable Systems

15) Tópicos Computacionais em Ciência de Dados

Ementa variável.

16) Tópicos Estatísticos em Ciência de Dados

Ementa variável.