

Ementa da Disciplina: Big Data for Finance Project

Curso: Ciência de Dados para Negócios

Informações Gerais

- **Curso:** Ciência de Dados para Negócios
- **Disciplina:** Big Data for Finance Project
- **Carga Horária:** 90 horas, sendo \approx 26 encontros de 4 horas/aula (19h às 22h30) cada.
- **Linguagens Utilizadas:** R ou Python

Descrição da Disciplina

A disciplina **Big Data for Finance Project** tem como objetivo capacitar os alunos a desenvolverem projetos aplicados à área financeira utilizando técnicas de análise de grandes volumes de dados (Big Data). Os alunos serão incentivados a trabalhar com linguagens R ou Python para a criação de modelos analíticos e de visualização, implementando soluções que poderão ser aplicadas em contextos reais. Ao final do curso, cada aluno terá criado uma página no formato de **GitHub Pages**, onde publicará seus projetos, compondo um portfólio individual voltado ao mercado de trabalho.

Objetivos

- Aplicar conceitos de Big Data em projetos financeiros reais;
- Desenvolver habilidades práticas em R ou Python para análise de dados financeiros;
- Criar soluções analíticas e de visualização de dados para problemas financeiros;
- Produzir um portfólio online no formato de posts em GitHub Pages.

Competências Desenvolvidas

- Manipulação e análise de grandes volumes de dados financeiros;
- Construção de modelos preditivos e diagnósticos;
- Visualização de dados para suporte à decisão;
- Comunicação e apresentação de resultados de forma profissional.

Conteúdo Programático

1. Criação de GitHub Pages com Quarto (quarto.org)

- Configuração inicial do Quarto
- Publicação de páginas no GitHub Pages

2. Geração de Dados Sintéticos

- Introdução aos dados sintéticos
- Métodos e ferramentas para criação de dados simulados

3. Geração de Features e Métricas Fundamentais para Séries Temporais Financeiras

- Construção de variáveis diferenciadas
- Técnicas para análise de estacionaridade

4. Forecasting de Séries de Preços na Teoria Econômica Básica

- Fundamentos de forecasting
- Aplicações práticas na previsão de preços

5. Teoria Básica de Markowitz e Modelagem Multiobjetivo

- Introdução ao modelo de Markowitz
- Implementação de um modelo biobjetivo para retorno e risco

6. Considerando o Perfil do Investidor na Tomada de Decisão

- Ajustes baseados no perfil do investidor
- Estratégias para decisões personalizadas

Cronograma de Aulas 2025

Data	Conteúdo Programado	Entrega Associada	Hrs/Aula	Obs
28/02/2025	Apresentação do curso e do projeto final. Configuração inicial do GitHub Pages e criação da primeira página usando Quarto. Formação de grupos para criação da GitHub Page colaborativa.	Criar uma GitHub Page simples seguindo o roteiro disponível em https://profrhozon.github.io/site/introportfoliogithubpages.html	4	
07/03/2025	Geração de Dados Sintéticos e Datasets Públicos: introdução a dados sintéticos e exploração de datasets reais para modelos de séries temporais.		4	

Data	Conteúdo Programado	Entrega Associada	Hrs/Aula	Obs
14/03/2025	Feature Engineering em séries temporais financeiras. Log-retornos e variâncias condicionais (GARCH).	Primeira página criada de cada grupo e apresentação dos grupos formados.	4	Sem entrega associada nesse dia.
21/03/2025	Técnicas de redução de dimensionalidade em séries temporais financeiras. O impacto do Big Data na decisão humana.	Post com o planejamento do projeto de cada grupo e apresentação da GitHub Page do grupo.	4	
28/03/2025	Modelos causais vs. modelos de Machine Learning (black-box) na previsão de séries temporais. Comparação de performance de modelos.	Post e apresentação da Análise Exploratória de Dados (AED) do conjunto de dados selecionado.	4	
04/04/2025	Simulação de cenários vs. forecasting. Estratégias de compra, venda e espera baseadas no perfil do investidor. Trade-off risco vs. retorno.	Post e apresentação do feature engineering dos dados financeiros.	4	
11/04/2025	Decisão Multicritério em Cenários Econômicos. Modelos de hierarquização e gestão da incerteza.	Composição da P1 (10 a 16 de Abril).	4	
18/04/2025	Feriado - Páscoa (Sem Aula)			
25/04/2025	Risco vs. Retorno vs. Incerteza. O impacto de eventos inesperados no mercado financeiro (The Black Swan).	Post apresentando previsões de retornos, volatilidades e janelas de oportunidade identificadas.	4	
02/05/2025	Introdução à Teoria de Alocação de Portfólio (Markowitz, Black-Litterman). Cálculo de retornos cumulativos dos ativos.	Post com matriz de decisão multicritério para cada portfólio.	4	
09/05/2025	Modelo de Markowitz: pesos dos ativos, fronteira eficiente, cálculo de risco e retornos esperados.	Post com valores projetados e métricas de retorno anualizados/mensalizados.	4	
16/05/2025	Efeitos da diversificação, entropia de Shannon, estatísticas do portfólio, avaliação de riscos (Sharpe Ratio, Sortino Ratio, Drawdown).	Post construindo a Fronteira Eficiente baseada nas previsões de retorno e risco.	4	
23/05/2025	Avaliação de Performance do Portfólio (Fama-French Factor Model). Benchmarks e fatores de risco.	Post sobre distribuições de retornos dos ativos (Histogramas vs. Kernel Density).	4	
30/05/2025	Drivers de Performance e risco do portfólio. Abordagem matricial e otimização orçamentária.	Post com indicadores de performance do portfólio (Sharpe, Sortino, Drawdown, assimetria, curtose).	4	
06/06/2025	Backtest da otimização de portfólio e comparação de estratégias.	Post sobre performance do modelo Fama-French.	4	Sem entrega associada.

Data	Conteúdo Programado	Entrega Associada	Hrs/Aula	Obs
13/06/2025	Teoria Moderna de Portfólio e Portfólio Ótimo: Sharpe Máximo, volatilidade mínima, Equal Weighted Risk.	Composição da P2 (12 a 18 de junho).	4	
20/06/2025	Feriado (Sem Aula)			
27/06/2025	Dinâmica de preços e inflação. Como a inflação afeta a precificação de ativos?	Post sobre portfólio otimizado vs. benchmark.	4	
04/07/2025	Prazo de entrega substitutiva (Sem Aula)			
11/07/2025	Recesso do Corpo Docente (Sem Aula)			
18/07/2025	Recesso do Corpo Docente (Sem Aula)			
25/07/2025	Sem Aula			
01/08/2025	Sem Aula			
08/08/2025	Algoritmos de Otimização Multiobjetivo para portfólio.	Post reflexivo sobre o impacto da inflação nos ativos do portfólio.	4	
15/08/2025	Métodos de Otimização Multiobjetivo para portfólios.		4	Post opcional.
22/08/2025	Modelos econométricos vs. Machine Learning para previsão de inflação.	Post sobre otimização do portfólio.	4	
29/08/2025	Impacto da inflação nos preços de ativos e commodities.	Post opcional sobre NSGA-II/III ou DEOptim aplicado aos dados.	4	
05/09/2025	Medidas de inflação e juros. Matemática financeira aplicada ao portfólio.	Post reflexivo sobre impacto da inflação no portfólio.	4	
12/09/2025	Revisão geral do projeto.	Sprint Review dos alunos com o professor.	4	
19/09/2025	Apresentação final dos projetos.	Composição da PFinal (vide calendário).	4	

Metodologia

A disciplina será ministrada de forma prática, com aulas expositivas, estudos de caso e exercícios em laboratório. Os alunos trabalharão em projetos individuais e em grupo, recebendo orientação para a publicação de seus resultados em GitHub Pages.

Critérios de Avaliação

- Participação e exercícios práticos: 30%
- Entrega de projetos intermediários: 30%
- Projeto final publicado no GitHub Pages: 40%

Referências recomendadas

- [1] Angrist, J. D., & Pischke, J. S. (2009). *Mostly Harmless Econometrics: An Empiricist's Companion*. Princeton University Press.
- [2] Angrist, J. D., & Pischke, J. S. (2014). *Mastering 'Metrics: The Path from Cause to Effect*. Princeton University Press.
- [3] Caetano, M. A. L. (2021). *Python e Mercado Financeiro: Programação Para Estudantes, Investidores e Analistas*. Editora Blucher. https://www.blucher.com.br/python-e-mercado-financeiro_9786555062403
- [4] Chambers, J. M. (2018). *Spark: The Definitive Guide: Big Data Processing Made Simple*.
- [5] **Skill Track: Big Data with R** (DataCamp): Programa de especialização em Big Data com R. Disponível em: <https://app.datacamp.com/learn/skill-tracks/big-data-with-r>.
- [6] DataCamp. (n.d.). *Financial Analytics in R*. Disponível em: <https://www.datacamp.com/courses/financial-analytics-in-r>
- [7] DataCamp. (n.d.). *Portfolio Analysis in R*. Disponível em: <https://www.datacamp.com/courses/portfolio-analysis-in-r>
- [8] **Big Data Fundamentals with PySpark** (DataCamp): Curso voltado à introdução aos conceitos fundamentais de Big Data utilizando PySpark. Disponível em: <https://app.datacamp.com/learn/courses/big-data-fundamentals-with-pyspark>.
- [9] DataCamp. (n.d.). *Python for Finance*. Disponível em: <https://www.datacamp.com/courses/python-for-finance>
- [10] DataCamp. (n.d.). *Quantitative Analyst with Python*. Disponível em: <https://www.datacamp.com/tracks/quantitative-analyst-with-python>
- [11] DataCamp. (n.d.). *Introduction to R for Finance*. Disponível em: <https://www.datacamp.com/courses/introduction-to-r-for-finance>
- [12] Gonzaga, S., & Freitas, W. (2020). *Tópicos de Finanças com R*. GitHub Pages. https://sillasgonzaga.github.io/curso_r_financas/
- [13] Gujarati, D. N., & Porter, D. C. (2009). *Basic Econometrics*. McGraw-Hill Education.
- [14] Hamberger, P. (2020). *Aprendendo R para Economia e Finanças*. GitHub Pages. https://focusfinance.github.io/focus_finance/
- [15] Hilpisch, Y. (2019). *Python for Finance: Mastering Data-Driven Finance*. O'Reilly Media. <https://www.amazon.com.br/Python-Finance-2e-Yves-Hilpisch/dp/1492024333>

-
- [16] James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). *An Introduction to Statistical Learning: with Applications in R*.
- [17] Lewinson, E. (2022). *Python for Finance Cookbook - Second Edition*. Packt Publishing. <https://loja.umlivro.com.br/python-for-finance-cookbook---second-edition-6628560/p>
- [18] Maddala, G. S. (2001). *Introduction to Econometrics*. Wiley.
- [19] *Mastering Econometrics with Joshua Angrist*. Disponível em: <https://www.masteringeconometrics.com/>
- [20] McKinney, W. (2017). *Python for Data Analysis*. O'Reilly Media.
- [21] Provost, F., & Fawcett, T. (2013). *Data Science for Business: What You Need to Know About Data Mining and Data-Analytic Thinking*.
- [22] Tatsat, H., Puri, S., & Lookabaugh, B. (2020). *Blueprints de Aprendizado de Máquina e Ciência de Dados Para Finanças: Desenvolvendo Desde Estratégias de Trades Até Robôs Advisors com Python*. Alta Books. <https://www.amazon.com.br/Blueprints-Aprendizado-M%C3%A1quina-Ci%C3%A4ncia-Finan%C3%A7as/dp/8550821829>
- [23] VanderPlas, J. (2016). *Python Data Science Handbook*. O'Reilly Media.
- [24] **Projeto VolPrev** (UFPR): Monografia de graduação de Rodrigo H. Ozon (2008) em Ciências Econômicas (UFPR). Disponível em: <http://www.leg.ufpr.br/doku.php/projetos:ehlers:volprev>.
- [25] Wickham, H., & Grolemund, G. (2017). *R for Data Science*. O'Reilly Media.
- [26] Wooldridge, J. M. (2010). *Econometric Analysis of Cross Section and Panel Data*. MIT Press.
- [27] Wooldridge, J. M. (2012). *Introductory Econometrics: A Modern Approach*. Cengage Learning.



Livros e artigos complementares indicados durante o curso. Algumas documentações de pacotes do R seriam interessantes e alguns conceitos teóricos serão apresentados gradualmente.



Sobre o Professor:

Rodrigo Hermont Ozon, economista apaixonado por econometria, pelas aplicações de modelos econômicos a problemas reais vivenciados na sociedade e na realidade das empresas. Atualmente professor da disciplina Big Data for Finance Project na FAE Business School.

Seus contatos podem ser acessados em:

-  GitHub
-  LinkedIn
- Lattes
- Ou então, me envie um e-mail

Para maiores detalhes e avisos, materiais e informações a respeito da ementa, programação e demais atualizações visite o site do prof. em
<https://profrhozon.github.io/site/>