

Simulando projetos de ML com dados e clientes reais em cursos de graduação

Qual é a opinião dos estudantes sobre isto?

INS^{PER}¹, PUC-SP²

Fabício Barth¹, Fabio Miranda^{1,2}, Marcelo Graglia²

Sessões Sprints no Curso de Ciência da Computação

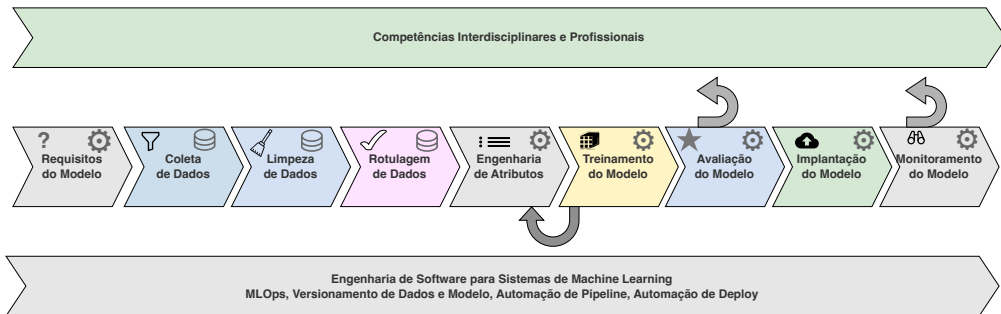
1	Developer Life				Ética, Computação e Sociedade
	Sessão Sprint 1 (embutida em DevLife)				
2	Projeto Ágil	Programação Eficaz	Bits e Processadores	Matemática do Contínuo	Ciência dos Dados
	Sessão Sprint 2 - Backend de Dados				
3	Arquitetura de Objetos e Times Ágeis	Técnicas de Programação	Álgebra Linear e Teoria da Informação	Lógica e Matemática Discreta	Inteligência Artificial e Robótica
	Sessão Sprint 3 - Inovação Social				
4	Projeto de Software e Gestão Ágil	Sistemas Hardware-Software	Algoritmos e Estruturas de Dados	Linguagens e Paradigmas	Machine Learning
	Sessão Sprint 4 - AI e Machine Learning				
5	Plataformas, Microserviços e APIs	Jogos e Interação	Análise de Algoritmos e Entrevistas Técnicas	Megadados	AI Startup
6	Cybersegurança	Projeto Final de Computação			
7	Eletiva	Eletiva	Eletiva	Estágio Supervisionado 1	
8	Eletiva	Eletiva	Eletiva	Estágio Supervisionado 2	

- Trilha de Times Ágeis e Projetos de Software Complexos
- Trilha de Inteligência Artificial e Machine Learning
- Trilha de Fundamentos e Matemática
- Trilha de Complexidade e Performance

- ▶ As sessões sprint acontecem no final de cada semestre, do 1º ao 4º.
- ▶ Nos sprints os alunos precisam apresentar soluções para problemas específicos.
- ▶ Cada sessão sprint tem duração de 3 semanas, com atividades em período integral.
- ▶ Cada sessão sprint tem objetivos de aprendizado distintos.
- ▶ As sessões sprint são projetadas para proporcionar uma experiência de aprendizado prática e aplicada.

Sessão sprint do 4º semestre

O **objetivo da sessão sprint** do quarto semestre é capacitar os alunos a desenvolver projetos **machine learning** de ponta a ponta, passando por **todas as etapas do ciclo de vida de um projeto real**, da definição do problema até a entrega de um sistema funcional.



Problemas tratados no sprint do 4º semestre

(2023-2) Identificar, com antecedência, os clientes com maior probabilidade de encerrar o serviço (**análise de churn**)

(2024-2) **Deteção da Deficiência** de Recombinação Homóloga (HRD) em amostras de câncer **a partir de dados de sequenciamento genético** de nova geração (NGS)

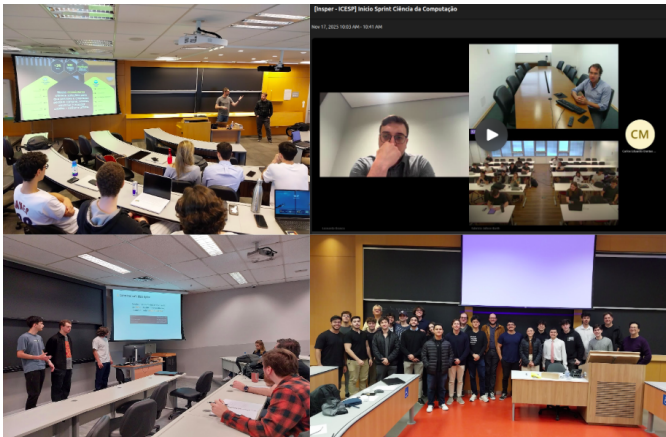
(2025-1) Identificação de **fraudes** em transações com cartão de crédito ou débito presenciais, sob o ponto de vista do adquirente

(2025-2) Construção automatizada de banco de dados clínicos a partir de **prontuários eletrônicos** desestruturados

(2026-1) **Precificação de imóveis**

The logo for anahealth, featuring the word "ana" in blue and "health" in a lighter blue, with a stylized human figure icon integrated into the "a" of "health".The logo for DASA genômica, with "DASA" in dark blue and "genômica" in pink, with a stylized DNA helix icon integrated into the "A" of "DASA".The logo for mercado livre, featuring a yellow circle with a blue handshake icon and the words "mercado livre" in blue.The logo for the Instituto do Câncer do Estado de São Paulo, featuring a green and blue stylized building icon and the text "INSTITUTO DO CANCER DO ESTADO DE SÃO PAULO" and "OCTAVIO FREITAS DE OLIVEIRA".The logo for QuintoAndar, featuring a purple square icon with a white "Q" and the word "QuintoAndar" in black.

Exemplos de interação com os clientes



Contrapartida das organizações parceiras: **definição do problema, disponibilização dos dados e tempo para reunião com os alunos.**

Rubrica utilizada

A rubrica utilizada em todos os semestres/projetos é similar. Os grandes grupos avaliativos são:

1. setup do projeto;
2. aquisição e pré-processamento dos dados;
3. engenharia de atributos;
4. desenvolvimento e avaliação dos modelos;
5. apresentação dos resultados;
6. deploy dos sistemas e modelos; e
7. organização da equipe.

Os objetivos da ementa e dos grupos avaliativos da rubrica são coerentes com os nove estágios de um projeto de machine learning.

Avaliação Qualitativa

Pergunta: *Descreva, em um parágrafo, como foi a sua experiência neste sprint, destacando os pontos positivos e negativos, bem como o que você exercitou e aprendeu durante o processo.*

Método de análise: codificação temática — respostas lidas, categorizadas e organizadas em temas representativos por **dois pesquisadores de forma independente**.

Semestres	Estudantes	Equipes	Observações
2024-2	29	7	Uma equipe foi composta por cinco integrantes e as demais por quatro.
2025-1	17	4	Uma equipe foi composta por cinco integrantes e as demais por quatro.
2025-2	28	7	Todas as equipes foram formadas por quatro integrantes.
Total	74	18	

Resultados

Aspectos positivos (recorrentes nas 3 turmas)

- ▶ Forte valorização do caráter prático e do uso de dados reais
- ▶ Interação com empresas parceiras
- ▶ Desenvolvimento técnico: CNN, TensorFlow, NLP, OCR, LLMs
- ▶ Práticas de MLOps: versionamento (DVC), pipelines e deploy (MLflow)

Desafios (recorrentes nas 3 turmas)

- ▶ Limitação de tempo (2025-1 e 2025-2)
- ▶ Complexidade e tratamento dos dados
- ▶ Versionamento de dados e uso do DVC
- ▶ Desorganização, tarefas imprevistas e trabalho em equipe

Síntese

Os sprints cumprem um papel relevante na formação dos estudantes ao proporcionar vivências próximas à realidade profissional, expondo-os não apenas às etapas técnicas do ciclo de vida de projetos de ML, mas também às incertezas, restrições de tempo, limitações de infraestrutura e desafios de comunicação típicos do contexto real.

Obrigado!



Artigo completo:
escaneie o QR code

Contato:
`fabriciojb@insper.edu.br`