



Pontifícia Universidade Católica do Paraná

Plano de Ensino

Escola/ Câmpus:	Escola Politécnica			
Curso:	Bacharelado em Ciência da Computação	Ano/Semestre:	2023/2	
Código/Nome da disciplina:	Data Science			
Carga Horária:	120 h/a			
Requisitos:	Não há.			
Créditos:	6 créditos	Período:	Turma:	Turno:
		6º	U	Noturno
Professor Responsável:	Jean Paul Barddal			

1. Ementa:

A disciplina de Data Science é de natureza teórico-prática e é ofertada para estudantes dos cursos de Bacharelado em Ciência da Computação. Ao longo do semestre, o estudante prepara, seleciona, analisa e desenvolve visualizações efetivas de dados, assim como aplica técnicas de redução de dimensionalidade e de tratamento de desbalanceamento entre classes, sempre objetivando melhorias nos processos de aprendizagem de máquina e de tomada de decisão. Ao final da disciplina, o estudante é capaz de atuar nas diversas etapas de projetos de preparação de dados, análise de dados e aprendizagem de máquina.

(Syllabus)

Data Science is a theoretical and practical course that is offered to Computer Science bachelor degree students. Throughout the semester, the student prepares, selects, analyzes, and tailors effective data visualizations, as well as applies dimensionality reduction and class imbalance handling techniques, always targeting improvements in machine learning and decision-making processes. At the end of the discipline, the student can work on different steps of projects that involve data handling, data analysis, and machine learning.

2. Relação com disciplinas precedentes e posteriores

Disciplinas anteriores

- Raciocínio Algorítmico
- Banco de Dados
- Resolução de Problemas Estruturados em Computação
- Métodos Quantitativos para Computação
- Big Data
- Aprendizagem de Máquina

Disciplinas posteriores

- Tomada de Decisão Usando Modelagem Matemática
- Avaliação de Desempenho de Sistemas

3. Temas de estudo

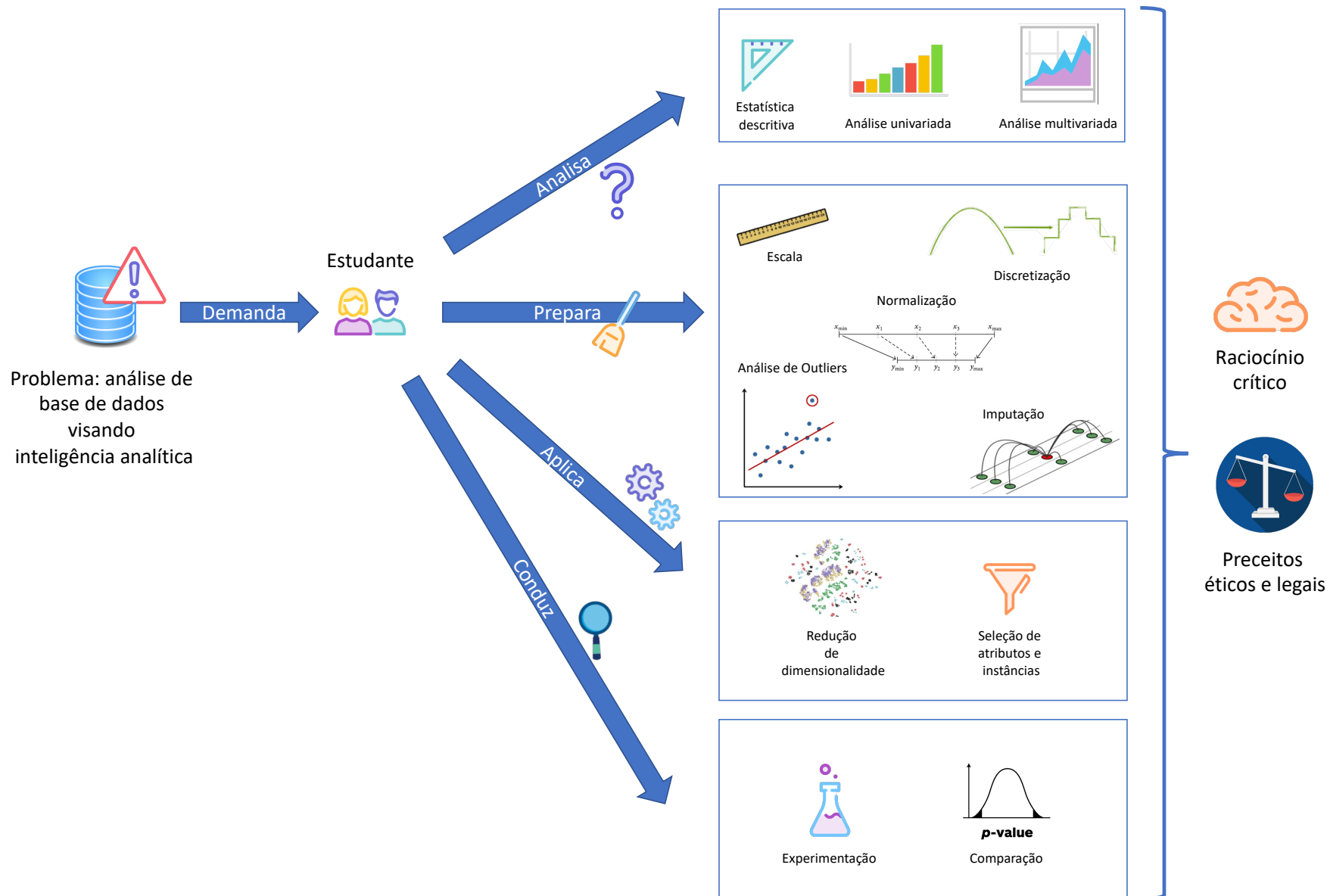
Abaixo são apresentados os principais temas da disciplina de Data Science, seguidos de seus respectivos tópicos teóricos e práticos.

- **Tema 1 (TE1): Análise exploratória de dados**
- **Tema 2 (TE2): Seleção de Dados**
- **Tema 3 (TE3): Visualização de Dados**
- **Tema 4 (TE4): Redução de Dimensionalidade**
- **Tema 5 (TE5): Análise de Desbalanceamento**
- **Tema 6 (TE6): Análise e comparação de modelos**

4. Resultados de Aprendizagem

Resultados de Aprendizagem	Temas de Estudo	Elemento de Competência (Internaliza, Mobiliza, Certifica) e Competência
Competência B – Desenvolver modelos computacionais inteligentes e eficazes aderentes a cenários diversos, seguindo princípios éticos.	TE1, TE2, TE3, TE4 e TE5	(Certifica) Elemento de competência B1 – Analisar o problema com base no pensamento computacional.
	TE1, TE2, TE3, TE4 e TE5	(Certifica) Elemento de competência B2 – Selecionar métodos computacionais aderentes ao contexto.
	TE2, TE3, TE4, TE5 e TE6	(Certifica) Elemento de competência B3 – Criar modelos computacionais inteligentes, seguindo princípios éticos.
	TE6	(Certifica) Elemento de competência B4 – Avaliar modelos computacionais.

5. Mapa Mental



6. Metodologia e Avaliação

Os resultados de aprendizagem desta disciplina serão trabalhados de acordo com a tabela apresentada abaixo. Nesta tabela, os resultados de aprendizagem (RAs) são relacionados com os seus respectivos indicadores de desempenho (IDs), os métodos utilizados para a apresentação conceitual e prática dos conteúdos, assim como os processos de avaliação via metodologias diagnósticas, formativas e somativas; a serem realizadas, sempre visando feedback rápido e efetivo para os estudantes.

Focando nas avaliações, estas podem ser diagnósticas, formativas ou somativas, conforme detalhamento abaixo:

- **Avaliações diagnósticas:** atividades com objetivo de acompanhar o aprendizado dos temas discutidos até o momento visando feedback imediato e de forma coletiva. A sugestão aqui é utilizar métodos como a linha da vida para engajar os estudantes a responder, mesmo que de forma não verbal, questões relacionadas ao conteúdo ministrado, permitindo ao professor elucidar conceitos que possivelmente estejam causando dúvidas e que os estudantes se auto avaliem e identifiquem tópicos a serem verificados antes das avaliações somativas.
- **Avaliações formativas:** atividades com enfoque prático que objetiva, por meio do trabalho em equipe, aplicar os conceitos ministrados pelo professor em problemas do mundo real. Estas avaliações permitem ao professor prover feedback aos alunos durante a própria aula, ainda de forma coletiva, mas focada em grupos menores de estudantes. Este tipo de atividade também permitirá que avaliação por pares seja realizada, onde equipes validem os trabalhos realizados pelas demais.
- **Avaliações somativas:** atividades com o objetivo de quantificar a performance dos estudantes, de forma individual, nos conceitos teóricos e práticos que envolvem a disciplina. Sugere-se aqui a aplicação de avaliações no formato de provas. O feedback para este tipo de atividade será realizado de forma colaborativa, no sentido de que o professor apresentará a resposta para todas as questões da avaliação para todos os estudantes, seguido por um feedback individual para cada estudante. Objetiva-se realizar a devolutiva destas avaliações em um prazo de uma a duas semanas após sua realização.

Alinhamento Construtivo			
Resultado de aprendizagem	Indicadores de desempenho	Processos de Avaliação	Métodos ou técnicas empregados**
Competência B – Desenvolver modelos computacionais inteligentes e eficazes aderentes a cenários diversos, seguindo princípios éticos.	B1: Analisar o problema com base no pensamento computacional. B2: Selecionar métodos computacionais aderentes ao contexto. B3: Criar modelos computacionais inteligentes, seguindo princípios éticos.	Avaliação diagnóstica: resolução de exercícios com cunho conceitual-teórico durante as aulas. <u>Feedback:</u> coletivo e imediato, durante as aulas. Avaliação formativa: resolução de exercícios com simulação realística. <u>Feedback:</u> coletivo em sala de aula. Avaliação somativa: trabalho de implementação de projeto realizado em equipes.	Métodos empregados: PBL, PjBL e Peer Instruction Meios de interação: Blackboard, Zoom, Mentimeter e Discord (usados de acordo com demanda e/ou instabilidade das opções anteriores)

	B4: Avaliar modelos computacionais.	<u>Feedback:</u> coletivo, para a equipe, com vistoria de código individualizada. <u>Recuperação:</u> ocorrerá no formato de trabalhos extra-classe, de acordo com a necessidade de cada estudante. Avaliação somativa: questões em provas escritas. <u>Feedback:</u> individual. <u>Recuperação:</u> ocorrerá na semana de recuperação.	
--	--	---	--

Para **aprovação** nesta disciplina, assume-se os seguintes critérios de aprovação:

- Caso o estudante **não** atinja a nota mínima esperada em um ou mais dos Ras (neste caso, competência a ser certificada), será oportunizada uma atividade de Recuperação, onde o estudante terá a oportunidade de recuperar os resultados não atingidos por meio de atividades específicas.
- Por possuir apenas um RA, a equação que representa o cálculo deste RA e da média semestral é representada, de forma algébrica, a seguir:

$$\text{Média semestral} = \frac{25\% * \text{Prova \#1} + 25\% * \text{Prova \#2} + 25\% * \textit{Checkpoint} + 25\% * \textit{Entrega Final}}{4}$$

- Ressalta-se que os critérios de avaliação das atividades, especialmente as de Checkpoint e Entrega Final, são apresentadas em documentos próprios disponíveis no Ambiente Virtual de Aprendizagem.

7. Cronograma de atividades

O cronograma previsto para a realização desta disciplina é fornecido abaixo. Mudanças no cronograma poderão ocorrer de acordo com o andamento da disciplina e servem para guiar o planejamento do curso.

Data	Atividades pedagógicas	Em aula / Pré-aula / Pós-aula / TDE	Carga horária
2/Agosto/2023	Apresentação do plano de ensino, ementa, formato de avaliação. Apresentação do TDE/Projeto. Contextualização sobre Data Science. Definição de equipes. Acolhida institucional.	Em aula	6h/a
16/Agosto/2023	Pandas e estatística descritiva de dados.	Em aula	6h/a
19/Agosto/2023 (reposição de 9/Agosto/2023)	Análise univariada e multivariada de dados.	Em aula	6h/a
23/Agosto/2023	Análise de correlações.	Em aula	6h/a
30/Agosto/2023	Visualização efetiva de dados.	Em aula	6h/a
20/Setembro/2023	Entrega do checkpoint. Revisão por pares do Checkpoint.	Em aula	6h/a
23/Setembro/2023 (reposição de 6/Setembro/2023)	Análise de Outliers.	Em aula	6h/a

27/Setembro/2023	Prova 01. (Aplicação a ser realizada por outro professor)	Em aula	6h/a
4/Outubro/2023	Semana acadêmica da Escola Politécnica. Feedback da prova 01.	Em aula	6h/a
11/Outubro/2023	Discretização, Escala, Normalização e Imputação. Seleção de atributos.	Em aula	6h/a
14/Outubro/2023 (reposição de 13/Setembro/2023)	Seleção de instâncias e o problema de desbalanceamento.	Em aula	6h/a
18/Outubro/2023	Testes de hipótese.	Em aula	6h/a
25/Outubro/2023	Prova 02.	Em aula	6h/a
1/Novembro/2023	Feedback da prova 02. Entrega de relatório e previsões.	Em aula	6h/a
8/Novembro/2023	Bancas de Projetos.	Em aula	6h/a
22/Novembro/2023	Recuperação.	Em aula	6h/a

29/Novembro/2023	Recuperação estendida, de acordo com a necessidade de cada estudante.	Em aula	6 h/a
------------------	---	---------	-------

8. Bibliografia

Básica:

de CASTRO, L. N., FERRARI, D. G. Introdução à mineração de dados: Conceitos básicos, algoritmos e aplicações. 2016.

FACELI, K., LORENA, A. C., GAMA, J., CARVALHO, A. C. P. de L. F. Inteligência Artificial: Uma Abordagem de Aprendizado de Máquina. 2011.

BONAFINI, C. F. Probabilidade e Estatística. Bibliografia Universitária Pearson. 1ª edição. 2015.

OLIVEIRA, F. E. M. Estatística e Probabilidade: com ênfase em exercícios propostos. LTC – Livros Técnicos e Científicos Editora Ltda. 3ª edição. 2017.

Complementar:

CHEN, Daniel Y. Pandas for everyone: Python data analysis. Addison-Wesley Professional, 2017.

KNAFLIC, Cole Nussbaumer. Storytelling with data: A data visualization guide for business professionals. John Wiley & Sons, 2015.

VANDERPLAS, Jake. Python data science handbook: Essential tools for working with data. "O'Reilly Media, Inc.", 2016.

CADY, Field. The data science handbook. John Wiley & Sons, 2017.

MAATEN, Laurens van der; HINTON, Geoffrey. Visualizing data using t-SNE. Journal of machine learning research, v. 9, n. Nov, p. 2579-2605, 2008.

CHAWLA, Nitesh V. et al. SMOTE: synthetic minority over-sampling technique. Journal of artificial intelligence research, v. 16, p. 321-357, 2002.

9. Adaptações

Nenhuma adaptação foi necessária.

10. Adaptações para práticas profissionais

Nenhuma adaptação foi necessária.