

UNIVERSIDADE DE SÃO PAULO
ESCOLA POLITÉCNICA

**Departamento de Engenharia de Computação e
Sistemas Digitais**

PCS 5787
Ciência dos Dados

Pós-Graduação em Engenharia Elétrica
3o. Quadrimestre de 2020

Apresentação do Professor

✓ Prof. Dr. Pedro Luiz Pizzigatti Corrêa

- E-mail: pedro.correa@usp.br
- Sala: C2-43;

● Professores colaboradores:

- Prof. Dr. Jorge Rady
 - E-mail: jorgerady@usp.br
- Profa. Dra. Jeaneth Machicao
 - E-mail: mj.machicao@gmail.com

● Secretaria de Pós-Graduação PPGEE

- E-mail: posgrad.eletrica.poli@usp.br

Fones: 3091-5733

<http://www.ppgEE.poli.usp.br/>

Tópicos Cobertos

- ❏ **Introdução:** Histórico e contexto de Ciência dos Dados. Domínio de aplicação, áreas de abrangência e aplicações voltados para a Ciência dos Dados ;
- ❏ **Seleção e definição de amostras:** técnicas estatísticas aplicadas à seleção e definição de dados a serem aplicados em experimentos computacionais.
- ❏ **Avaliação de modelos computacionais:**
- ❏ **Modelos de gestão de dados**
- ❏ **Escalabilidade e manipulação de dados:**
- ❏ **Análise de dados e Visualização de Dados;**
- ❏ **Tendências e temas de pesquisa na área:** Análise de dados voltada para identificação de padrões. Análise de comportamento em séries temporais.

Método de Ensino

- **Aulas Teóricas:** o tema de aula será desenvolvido a partir de leitura prévia de textos básicos sobre o assunto tratado. Durante o curso serão distribuídas listas de exercícios com o objetivo de fixar os assuntos abordados. Durante a disciplina será desenvolvido um projeto prático, que será documentado num formato de artigo e apresentado num seminário.
- **Aulas práticas:** Parte das aulas serão práticas e poderá ser utilizada para resolução dos exercícios.
- **Atividades extraclasse:** leitura de artigos e outras referências para cada aula. Resolução de lista de exercícios. Preparação do projeto de artigo. Preparação de Seminário. Participação em Seminários e Workshops recomendados que ocorrerão no período da disciplina.
- **Projeto:** será desenvolvido, durante a disciplina, um projeto a ser apresentado num seminário e documentado num formato de artigo. O projeto deverá ser em duplas/trios (definidos pelo Professor).
- **Exercício:** desenvolvido durante o transcorrer da disciplina, relacionado a temas da disciplina

Avaliação

➤ Critério de Avaliação:

Método: Exercícios, Projeto e Participação:

- Exercícios: após a conclusão de cada tópico do programa (individual);
- Projeto: Artigo/Seminário: elaborado em duas etapas (Resumo, Completo) (dupla/trio);
- Participação: participação nas aulas e em atividades extraclasses.

Critério: Exercícios: 40%; Projeto: 50%; Participação: 10%.

Planejamento das aulas

<u>Aula</u>	<u>Data</u>	<u>Agenda</u>
1	17/09	Apresentação da Disciplina. Orientação sobre a preparação dos projetos: Prof. Pedro, Prof. Jorge e Profa. Jeaneth
2	24/09	Introdução e Modelos de Gestão de Dados. Lista de Exercícios 1: Prof. Pedro
3	01/10	Seleção e definição de amostras: Prof. Pedro – Lista de Exercícios 2: Prof. Pedro – Lista de Exercícios 2 Definição dos Grupos Prof. Pedro
4	08/10	Avaliação de modelos computacionais. Lista de Exercícios 3: Prof. Pedro
5	14/10	WORKSHOP com o IBGE – Colaboração com o IBGE e apoio projeto PARSEC (FAPESP)
	15/10	Modelos Computacionais para Estimadores – Deep Learning – Profa. Dra. Marina Jeaneth Entrega da Proposta de Projeto (Grupos)
6	22/10	Modelos Computacionais para Estimadores – Deep Learning – Profa. Dra. Marina Jeaneth – Lista de Exercícios 4.
7	29/10	Infraestrutura computacional para processamento e armazenamento de dados. HPC - Cluster/Grid/Cloud – Lista de Exercícios 5.
8	05/11	Infraestrutura computacional para processamento e armazenamento de dados. HPC/Cluster/Grid/Cloud –Lista de Exercício 6.
09	12/11	Análise e Visualização de Dados. Prof. Fernando (ESALQ/USP) - Lista de Exercícios 7
10	19/11	Análise e Visualização de Dados. Lista de exercícios 8. Prof. Pedro (ESALQ/USP)
11	26/11	Seminário Entrega/“Submissão” (draft) do artigo.
12	03/12	Seminário para apresentação dos artigos. Participação de todos.
13	10/12	Seminário para apresentação dos artigos. Participação de todos. Prazo final para entrega dos artigos revisados (e-disciplinas).

Referências

- As referências bibliográficas serão disponibilizadas ao final das aulas.
- Material da Disciplina:
<https://edisciplinas.usp.br/>
- Grupo de Estudos, Pesquisa e Extensão em Big Data do PPGEE/EPUSP - wds.poli.usp.br

Preparação para a próxima aula

Introdução a Ciência dos Dados e Big Data

1) Assistir as palestras do I, III e V Workshop on Data Science (POLI/USP):
(wds.poli.usp.br/wds5)

<http://iptv.usp.br/portal/video.action?idItem=33468>

<https://www.youtube.com/watch?v=A4kqJ4RrIlc&feature=youtu.be>

<https://www.youtube.com/watch?v=ehYLzYc2DxM&feature=youtu.be>

<https://www.youtube.com/watch?v=PR9498qwVyc&feature=youtu.be>

VWDS - 18 February 18th: <https://www.youtube.com/watch?v=Aos63aulFO4>

2) Responder as questões:

- a) O que é Data Science e Ciclo de Vida dos Dados. Quais são os desafios citados na área de Data Science.
- b) Dê exemplo de aplicações de visualização de Dados;
- c) O que são os projetos DataONE, ARM e GoAmazon – Por que é um exemplo de Aplicação de Ciência dos Dados – Big Data ?
- d) Preparar uma proposta de Projeto para a Disciplina (Meia página, destacando o assunto e datasets envolvidos)
- e) Incluir informações no Relatório sobre o tema da pesquisa na pós-graduação, programa de pós-graduação e nome do orientador.
- f) Entregar no e-disciplinas até 19/09 (sábado) – e-disciplinas