

**UNIVERSIDADE DE SÃO PAULO
ESCOLA POLITÉCNICA**

PTC 3467

Ciência dos Dados

(Introdução aos Algoritmos em Automação)

Graduação em Engenharia Elétrica

2o. Semestre de 2020

Docentes Responsáveis

📁 Prof. Pedro Luiz Pizzigatti Corrêa

✓ e-mail: pedro.correa@usp.br



📁 Prof. Felipe Miguel Pait

✓ e-mail: pait@usp.br



📁 Prof. Fuad Kassab Jr.

✓ e-mail: fuad@lac.usp.br

📁 Prof. José Reinaldo Silva

✓ e-mail: reinaldo@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:** Serão ministradas aulas para resolução dos exercícios e Workshop IBGE (uso de notebook pessoal) e acesso a infraestrutura de nuvem da disciplina – AWS.
- **Atividades extraclasse:** leitura de artigos e outras referências para cada aula. Resolução de lista de exercícios, quando não concluídas em aula. Preparação do projeto. 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 duas ou quatro pessoas (máximo) a ser definido pelos professores.
- **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) (grupo 3-4);
- Provas: 2 provas. Média deve ser ≥ 5 .
- Participação: discussão nas aulas, participação nos Fóruns – apoio aos colegas na resolução dos exercícios e sobre uso de ferramentas.

Critério: Exercícios: 30%; Projeto: 40%; Prova: 20%; Participação: 10%.

Planejamento das aulas

Semana	Data	Assunto
1	19/ago	Apresentação da Disciplina. Projeto da disciplina - Introdução à Ciência dos Dados. Prof. Pedro (EPUSP)
2	26/ago	Introdução e Modelos de Gestão de Dados. Orientação sobre a preparação dos projetos. Lista de Exercícios 1. Palestra 1 (10h-11h) Prof. Pedro (EPUSP) – Definição dos Grupos
3	02/set	Seleção e definição de amostras. Palestra 2 (10-11h). Prof. Pedro (EPUSP).
4	09/set	Avaliação de modelos computacionais. Lista de Exercícios 2. Envio da Proposta do Projeto. Prof. Pedro
6	16/set	Infraestrutura de <i>Cloud</i> para processamento e análise de dados. Lista de Exercícios 3. Prof. Reinaldo (EPUSP)
5	23/set	Workshop IBGE. Prof. Pedro (EPUSP).
8	30/set	Prova 1. Palestra 3 – Modelos computacionais - Prof. Pedro (EPUSP). Entrega do Planejamento do Projeto
9	07/out	Técnicas de modelagem computacional. Profa. Dra. Marina Jeaneth Machicao (EPUSP)
10	14/out	Técnica de modelagem computacional <i>Deep Learning</i> . Lista de Exercícios 4. Profa. Dra. Marina Jeaneth Machicao (EPUSP)
11	21/out	Análise e Visualização de Dados. Lista de exercícios 5– Prof. Pedro (EPUSP) Apresentação dos resultados parciais do projeto
12	28/out	Feriado – Não haverá aula – Dia do Funcionário Público
13	04/nov	Infraestrutura computacional para processamento e armazenamento de dados. Arquitetura de <i>Big Data</i> e Modelagem de Dados . Lista de exercícios 4 (Prof. Pedro)
14	11/nov	Hands on Banco de Dados Não Relacionais. Lista de exercícios 5. – Palestra 4. Prof. Pedro (EPUSP)
15	18/nov	Desenvolvimento do projeto. Prof. Pedro (EPUSP)
16	25/nov	Prova 2. Palestra 5. Prof. Pedro (EPUSP)
17	02/dez	Apresentação dos Projetos. Profs. Pedro/Felipe/Fuad/Reinaldo
18	09/dez	Apresentação dos Projetos. Profs. Pedro/Felipe/Fuad/Reinaldo/
19	16/dez	Apresentação dos Projetos. 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

Próxima Aula

● Introdução a Ciência dos Dados e Big Data:

1) Assistir as palestras I e II Workshop on Data Science

(<http://wds.poli.usp.br/wds/>):

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

<https://www.youtube.com/watch?v=J4JQ-lHteB0&t=2288s>

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 são exemplos de Aplicações de Ciência dos Dados – Big Data ?
- d) Preparar ideias de temas sobre proposta de Projeto para a Disciplina (Tema, justificativa, disponibilidade de dados abertos);
- e) Incluir informações no Relatório: projetos anteriores relacionados (estágios/IC), cursos, linguagens conhecidas, té(Python, R, C++, etc).
- f) Entregar o documento com as respostas hoje (19/08) até 23h59min