**Introdução aos Algoritmos em Automação – PTC 3467**

**Assunto: Ciência dos Dados**

**Profs. Pedro Luiz Pizzigatti Corrêa, Felipe Miguel Pait, Fuad Kassab Jr. e José Reinaldo Silva**

**Objetivo**

Apresentar tópicos relacionados a estudos para descoberta de conhecimento e predição em conjunto de dados, abordando a multidisciplinariedade do tema. Contextualizar e desenvolver a temática de Ciências dos Dados aplicando metodologias para lidar com grandes bancos de dados (*Big Data*), análise de dados multidimensional, processamento e visualização de dados espaço-temporal.

## Programa Resumido

Introdução e histórico de Ciência dos Dados e modelos de gestão de dados. Técnicas para definição e seleção de amostragens. Avaliação de Modelos computacionais. Escalabilidade e manipulação de grandes massas de dados, banco de dados relacional e não relacional. Análise de dados com métodos estatísticos, análise de dados com técnicas de aprendizado de máquina. Conceitos e métodos para a gestão de dados de *Big Data*, para aquisição, integração, curadoria visualização e apresentação de dados multidimensionais e espaço-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 práticas. Será necessário uso de notebook pessoal. Será disponibilizada infraestrutura de *cloud* AWS.

**Atividades extraclasse**: leitura de artigos e outras referências para cada aula. Resolução de lista de exercícios. Preparação do projeto de artigo. Preparação de Seminário.

**Projeto**: será desenvolvido, durante a disciplina, um projeto a ser apresentado num seminário e documentado num formato de artigo.

**Exercício:** desenvolvido durante o transcorrer da disciplina, relacionado a temas da disciplina.

## Avaliação da Aprendizagem

**Método:** Exercícios, Projeto, Prova e Participação.

* Exercícios: após a conclusão de cada tópico do programa (individual);
* Projeto: Artigo/Seminário: elaborado em duas etapas (Resumo, Completo);
* Prova: provas a serem respondidas individualmente
* Participação: participação nas aulas via chat e discussões. Participação nos Fóruns para resolução dos exercícios e apoio aos colegas em atividades extraclasse.

**Critério:** Exercícios: 30%; Projeto: 40%; Prova 20%; Participação: 10%. Se Média da Prova (MP) < 5, avaliação final = MP.

## Horário: 7h30min-11h00min (quarta-feira)

## 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 EPUSP - [wds.poli.usp.br](file:///Users/Pedro1/Documents/Textos/Pedro/poli/Po%CC%81s%20Graduac%CC%A7a%CC%83o/PCS%20-%20Data%20Science/Planejamento/wds.poli.usp.br)

|  |  |  |
| --- | --- | --- |
| **Semana** | **Data** | **Assunto** |
| 1 | 19/ago | Apresentação 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)  |
| 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. **Envio da Proposta do Projeto.**  Prof. Pedro |
| 6 | 16/set | Lista de Exercícios. **Envio da Proposta do Projeto.**  Prof. Pedro |
| 5 | 23/set | Modelagem de Dados - Acesso a Data Sets - Exercícios |
| 8 | 30/set | **Prova 1**. Palestra 3 – Modelos computacionais - Profs. Pedro (EPUSP). |
| 9 | 07/out | Preparação do Data Management Planning e Canvas do Projeto |
| 10 | 14/out | Workshop IBGE. Prof.. Pedro (EPUSP). |
| 11 | 21/out | Técnicas de modelagem computacional. Prof. Dra. Marina Jeaneth Machicao (EPUSP) |
| 12 | 28/out | Infraestrutura computacional para processamento e armazenamento de dados em Nuvem.. (Prof. Reinaldo) |
| 13 | 04/nov | Técnica de modelagem computacional Deep Learning. Lista de Exercícios. Prof. Dra. Marina Jeaneth Machicao (EPUSP) |
| 14 | 11/nov | Desenvolvimento do projeto. Prof. Pedro (EPUSP) – **Entrega do Projeto – Caracterização e Preparação dos Dados** |
| 15 | 18/nov | Análise e Visualização de Dados. Lista de exercícios – 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 entrega do relatório finai (e-disciplinas).** |