

Análise de texto de fontes desestruturadas e WEB

Maciel Calebe Vidal

macielcv@insper.edu.br

Objetivos de aprendizagem

- **Extrair dados** de arquivos **desestruturados** como PDFs, imagens e páginas Web
- **Identificar informações relevantes** em corpus textuais
- Utilizar técnicas de **machine learning** para **análise dos dados** extraídos.

Aulas

- Semanais
- Qui 18:00 às 20:00

Atendimento

- A cada duas semanas (Ver agenda do aluno)
- Seg 09h45 às 11h15
- Remoto
 - Podemos alterar para presencial
- Resolução de dúvidas

Aulas - Dinâmica

- Parte teórica
- Parte prática
- Participativa!

Avaliações

- Projeto (**PRT**)
 - Em equipe
 - Enunciado disponibilizado ao longo do semestre
 - Tem que ser um projeto novo
- Entregas de atividades (**ATV**)
 - Código
 - Notebooks
 - Prazo: geralmente até a próxima aula
 - Não atrase!

Nota Final

- Se **PRT** ≥ 4 e **ATV** ≥ 5 :

$$\mathbf{NF} = 0,5\mathbf{PRT} + 0,5\mathbf{ATV}$$

Senão

$$\mathbf{NF} = \min(\mathbf{PRT}, \mathbf{ATV})$$

- Será utilizada a tabela oficial de conversão de conceito para nota do Blackboard
 - I=0, D=4, C=5, C+=6, B=7, B+=8, A=9, A+=10

Auxílio IA

- Pode usar tudo!
 - Idealmente, itere até produzir uma resposta bacana
 - Escreva (no final das atividades) onde, como e quais ferramentas utilizou (você não será penalizado)

Código de ética

- MultiInsper
- Consultou algo no Stack Overflow?
 - Adicione o link!
 - Tudo bem copiar um trecho pronto que te ajude a resolver uma parte de um problema
 - Você não deve copiar a resposta pronta de um exercício
- Seja legal! Dê crédito aos seus colegas caso utilize parte de projetos dos semestres anteriores



Dúvidas?

Obrigado pela participação!