

Enunciado T2

No primeiro trabalho, o foco foi analisar sistemas de aprendizado de máquina que trabalhassem com dados estruturados. O trabalho T2 consistirá em elaborar um projeto que trabalhe com dados não estruturados. Existem duas temáticas possíveis de trabalho: Processamento de Linguagem Natural, ou Visão Computacional. Ideias alternativas podem ser propostas e debatidas com o professor.

Espera que o aluno:

- Identifique um problema que pode ser solucionado com aprendizado de máquina
- Encontre e utilize um dataset que seja válido para o problema proposto
- Realize o pré-processamento e análise necessárias
- Aplique um ou mais dos algoritmos vistos em aula para resolver o problema
- Pesquise e utilize as métricas apropriadas para poder avaliar a solução proposta
- Domine os conceitos utilizados
- Analise a solução proposta, traga exemplos de casos em que a solução possui facilidade para resolver e casos em que a solução possua dificuldade
- Demonstre o algoritmo funcionando durante a etapa de perguntas e respostas (DEMO)

Como ideias de aplicações em visão computacional: Identificação de rostos sem máscara, identificação de dígitos escritos a mão, identificação de formas (triângulos, círculos e quadrados) desenhados a mão, identificação de animais em imagens etc.

Como ideias para aplicações de NLP: Análise de sentimento em textos de redes sociais, tradução de máquina, identificação de entidades nomeadas, correção ortográfica automática, sumarização de texto, identificação de fake News etc.

Entrega:

- Número de Integrantes: de 2 a 6.
- Data de Entrega: 26/11.
- Formato da Entrega: Link/arquivos no Moodle e Apresentação em Sala de Aula
- Entregáveis:
 - a. Arquivo README.txt com nome e matrícula dos integrantes, e descrição dos procedimentos necessários para que o professor replique os resultados;
 - b. Código fonte com toda implementação (pode constar na entrega o fonte ou um link para repositório github, gitlab,...);
 - c. Relatório técnico em formato PDF/PPT contemplando:

- I. Referencial Teórico Sucinto com as tecnologias e referenciais teóricos utilizados.
- II. Delimitação do Desenvolvimento: o que foi aproveitado de código fonte de repositórios públicos e o que foi implementado pelo grupo. É necessário deixar essa distinção bem especificada.
- III. Quais os procedimentos realizados pelo grupo para coletar e pré-processar os dados, assim como para treinar e avaliar o modelo.
- IV. Resultados e Conclusão, explicando os resultados atingidos pelo grupo e possíveis melhorias que poderiam ser implementadas no trabalho.
- V. Referências: Bibliografia utilizada.