

Descrição

O trabalho final da disciplina consiste em uma apresentação a ser realizada para a turma. O objetivo deste trabalho é exercitar conceitos de interesse dos alunos e fomentar a discussão entre a turma em relação aos mesmos. O trabalho deve ter conexão com a disciplina de Aprendizado Não Supervisionado, sejam conteúdos vistos ou não vistos em aula. Os grupos podem optar por uma dentre as seguintes opções:

Opção 1: Aplicação prática de Aprendizado Não Supervisionado

O grupo deve buscar uma base de dados existente, formular hipóteses a serem investigadas e realizar experimentos aplicando conhecimentos de Aprendizado Não Supervisionado. O objetivo desta opção é exercitar de forma mais prática o aprendizado no curso, bem como estimular a discussão de diferentes abordagens em relação a problemas concretos.

Pontos importantes

1. A base de dados escolhida não pode ser sintética e deve exemplificar algum problema real a ser resolvido (ou compreendido). Alguns lugares para procurar por bases de dados:
 - [Google Dataset Search](#)
 - [Kaggle Datasets](#)
 - [Knoema](#)
 - [Data Portal](#)
 - Dados de diferentes países e cidades, como por exemplo [Analyze Boston](#), [Data.gov \(EUA\)](#) e [UK Data Service](#)
2. Deve ser desenvolvido um trabalho de ciência de dados, englobando tarefas como análise exploratória de dados, elaboração de hipóteses (e.g. o que se deseja avaliar/experimentar), preparação e pré-processamento, aplicação de algoritmos de aprendizado não supervisionado (e supervisionado, caso julgue-se interessante), entre outros.
3. A apresentação deve englobar, no mínimo, descrição da base de dados, análise e preparação dos dados, algoritmos utilizados, resultados obtidos (com utilização de métricas pertinentes) e conclusões sobre os resultados. Deve-se mostrar um Jupyter Notebook com os passos realizados.

Opção 2: Algoritmos e/ou técnicas de Aprendizado Não Supervisionado

O grupo deve escolher, estudar e compreender um algoritmo ou técnica de Aprendizado Não Supervisionado e apresentar para a turma. O objetivo desta opção é expandir o que foi visto em aula através de áreas de interesse que utilizem técnicas de Aprendizado Não Supervisionado. Algumas sugestões de assuntos:

- Áreas e Aplicações: Detecção de Anomalias, Regras de Associação, Topic Model, Visão Computacional, Processamento de Linguagem Natural

- Algoritmos: Self-Organizing Maps, BIRCH, Spectral, Clustering, Mean Shift, Random Projection, Independent Component Analysis, Transformers

Pontos importantes

1. Independentemente de ser uma área, aplicação ou um algoritmo, é importante introduzir o tema e o contexto, fazer uma conexão com o que foi visto em aula e apresentar um exemplo prático do algoritmo funcionando.
2. O caso prático pode ser um exemplo pronto existente, desde que o grupo explique o funcionamento com suas palavras e seu entendimento. Códigos de exemplo serão disponibilizados para a turma.

Opção 3: Ferramentas, frameworks, bibliotecas ou Web APIs de Machine Learning

Como uma terceira opção, o grupo pode focar mais em uma solução do ponto de vista tecnológico, podendo explorar ferramentas, frameworks, bibliotecas ou Web APIs de Machine Learning. O objetivo desta opção é poder explorar casos como a aplicação de K-means em um contexto distribuído através de Spark ou como colocar pipelines de treinamento de dados em produção com soluções como Dataiku. O foco maior é no aspecto técnico.

Pontos importantes

1. A tecnologia utilizada deve ser bem explicada, mostrando o funcionamento passo-a-passo. Ao utilizar uma determinada ferramenta, é necessário explicar detalhes como origem, para o que serve, que outras opções existem, entre outros.
2. Assim como na Opção 2, o caso prático pode ser um exemplo pronto existente, desde que o grupo explique o funcionamento com suas palavras e seu entendimento. Códigos de exemplo serão disponibilizados para a turma.

Informações Gerais e Avaliação

- Os alunos deverão se organizar em **6 grupos** diferentes, tendo 5 componentes cada um (provavelmente teremos um ou dois grupos com 4 alunos).
- Após formar os grupos, os alunos devem escolher uma das opções e **validar com o professor antes de iniciar o trabalho**.
 - O professor estará a disposição para ajudar a validar materiais e referências encontradas assim como propostas de apresentação. Aproveite para começar a organização do trabalho cedo!
- Os grupos terão 20 minutos para apresentação e mais 5 minutos para perguntas/discussões.
- Independentemente da opção escolhida, todos os trabalhos devem ter slides com explicações e código.
- Critérios de avaliação sobre a apresentação
 - Explicações claras, com suas palavras
 - Conhecimento em relação ao material
 - Exemplo prático/código demonstrado
- Critérios de avaliação sobre o material
 - Qualidade geral dos slides

- Organização do conteúdo (narrativa da apresentação)
- No geral, deve ajudar os colegas a entenderem o assunto (ter fluxos com passo-a-passo, por exemplo)
- Aproveite a oportunidade para explorar um assunto de seu interesse ou algo pertinente para o trabalho final de curso 😊.

Entrega Final

- Slides em PDF ou PPT (Power Point), colocados no Moodle em atividade que será criada.
- Jupyter notebook com os experimentos realizados (.ipynb e um .html ou .pdf) ou código fonte em uma pasta compactada (.zip).