

# Ciência de Dados II

3o Tech AD



# Ciência de Dados II



## Machine Learning – Aprendizado Não Supervisionado

- Clusterização: K-Means, Hierarchical Clustering, DBSCAN
- Otimização: Redução de dimensionalidade
- Avaliação de clusterização
- Aplicações práticas: segmentação de clientes, agrupamento de documentos



# Ciência de Dados II

## Técnicas de Redução de Dimensionalidade

- PCA (Principal Component Analysis)
- t-SNE (t-Distributed Stochastic Neighbor Embedding)
- UMAP (Uniform Manifold Approximation and Projection)
- Autoencoders (Redução de dimensionalidade com Redes Neurais)



# Ciência de Dados II

Redes Neurais com TensorFlow/Keras e PyTorch

- Estrutura básica de redes neurais
- Implementação em TensorFlow/Keras e PyTorch
- Técnicas avançadas: regularização, otimização de hiperparâmetros
- Transfer learning e fine-tuning de modelos pré-treinados

# Ciência de Dados II



## Processamento de Linguagem Natural (NLP)

- Tokenização, stemming e lematização
- Vetorização de texto: Bag of Words, TF-IDF
- Modelos baseados em embeddings: Word2Vec, GloVe
- Introdução a transformers e BERT



# Ciência de Dados II

## Automação e Deployment de Modelos

- Automação de experimentos com MLFlow
- Registro, rastreamento e reprodutibilidade de modelos
- Deployment via APIs REST e TensorFlow Serving
- Desenvolvimento de APIs para modelos de IA

# Ciência de Dados II



Projeto Final – desejável!

- Definição do problema e planejamento da solução
- Implementação de um pipeline completo de Ciência de Dados
- Treinamento, validação e deployment
- Apresentação e análise dos resultados

