



## Treinamento CIS - 4º Período

### 1. Convolutional Neural Networks (CNN)

- a. Redes Neurais Convolucionais:
  - i. Pré-Processamento de Imagens;
  - ii. Dropout;
  - iii. Pooling;
  - iv. BatchNormalization;
  - v. Callbacks;
  - vi. Transfer Learning;
  - vii. Freezing Layers.
- b. Manipulação de Diretórios com bibliotecas;
- c. Técnicas de avaliação e visualização do treinamento;
- d. Implementação usando TensorFlow.

### 2. Natural Language Processing (NLP)

- a. Pré-processamento de textos:
  - i. Lowercase, remoção de pontuação, etc;
  - ii. Remoção de stopwords;
  - iii. Stemming/Lemmatization;
- b. Representação vetorial de textos:
  - i. Bag-of-words (BOW);
  - ii. Term-frequency inverse document-frequency (TFIDF);
- c. Modelos de classificação:
  - i. Modelos clássicos de Machine Learning (Naive Bayes, SVM, etc);
  - ii. Redes Neurais Recorrentes (RNNs);
  - iii. Long-short-term memory (LSTM).

### 3. Conteúdos adicionais

- a. Pré processamento
  - Part-of-speech (POS) tagging;
  - Named entity recognition (NER)
  - Tokenização, vetorização e n-grams.
- b. Representação vetorial de textos
  - Word embeddings
- c. Modelos de classificação
  - Gated recurrent unit (GRU)

### 4. Conteúdos de apoio

- a. CNN
  - i. [Playlist Neural Networks 3blue1brown](#) - Playlist que explica o que são Redes Neurais, Gradiente descendente e Backpropagation;
  - ii. [Convolutional Neural Networks](#) - Aula do MIT sobre fundamentos do Deep Learning;



- iii. [Curso Redes Neurais Convolucionais](#) - Aulas do curso 4 da especialização em Deep Learning do deeplearning.ai;
- iv. [DeepLearning Book](#) - Livro em português sobre Deep Learning. Para o quarto período, recomenda-se os capítulos 40 ao 47;
- v. [Hands-on Machine Learning with Scikit-Learn, Keras, and TensorFlow](#) - Livro completo: Para o quarto período (CNN) recomenda-se o capítulo 11, 12 e 14.

**b. NLP**

- i. [NLP Tensorflow - deeplearning.ai](#) - Curso do Coursera ensinando a aplicar RNNs com tensorflow;
- ii. [Vídeo MIT RNN](#) - Aula do MIT sobre RNNs no youtube;
- iii. [Hands-on Machine Learning with Scikit-Learn, Keras, and TensorFlow](#) - Livro completo: Para o quarto período (NLP) recomenda-se os capítulos 13 e 16.

**5. Tarefas**

**a. CNN**

- i. Base de dados sugerida para o período - [Classificação - Cachorros, Gatos e Pandas](#)
- ii. Fazer função para preparar os diretórios dos arquivos para o formato necessário para o TensorFlow (do método `flow_from_directory`);
- iii. Fazer uma rede convolucional do zero em TensorFlow e avaliar no dataset;
- iv. Pegar uma rede pré treinada (Transfer Learning) e avaliar no dataset, comparando diferenças com o método anterior;
- v. Avaliar métodos de regularização e data augmentation;
- vi. As atividades descritas são apenas sugestões e podem ser modificadas;
- vii. No meio do período, haverá uma reunião com o monitor para consolidação das informações;
- viii. A entrega é individual e deverá ser colocada no seu GitHub pessoal

**b. NLP**

- i. Base de dados sugerida para o Período - [Amazon Musical Instruments Reviews](#);
- ii. Fazer uma função de pré-processamento dos dados usando SpaCy, NLTK e/ou RegEx (lowercasing, remover pontuações, stopwords, lemmatize, etc);
- iii. Tokenizar os textos com os métodos de BOW e TFIDF usando o sklearn;



- iv. Treinar um ou mais classificadores usando métodos clássicos de Machine Learning (Naive Bayes, SVM, etc);
- v. Fazer a tokenização dos textos usando o Tensorflow (função Tokenizer e método text\_to\_sequences);
- vi. Treinar uma RNN simples do zero usando Tensorflow com Embedding layer;
- vii. Substituir a RNN simples por uma LSTM unidirecional. Em seguida, por uma bidirecional, e comparar os resultados;
- viii. As atividades descritas são apenas sugestões e podem ser modificadas;