



## Treinamento CIS - 2º Período

### 1. Conteúdos do Período

- a. Limpeza e manipulação de dados;
- b. Algoritmos clássicos de Machine Learning:
  - i. KNN (K - Nearest Neighbours);
  - ii. Support Vector Machine;
  - iii. Decision Tree;
  - iv. Random Forest;
  - v. Naive Bayes
- c. Oversampling e Undersampling;
- d. Classificação binária;
- e. Classificação multiclasse;
- f. Métricas de Avaliação:
  - i. Cross-Validation;
  - ii. Confusion Matrix;
  - iii. Precision e Recall.
- g. Ensemble Learning.

### 2. Conteúdo de apoio

- a. [Machine Learning Full Course for Beginners](#) - Vídeo com as bases do Machine Learning e explicação de alguns algoritmos.
- b. [Machine Learning Sentdex](#) - Playlist com diversos assuntos de Machine Learning. Recomenda-se os que tratam dos conteúdos do período
- c. [Machine Learning Fundamentals StatQuest](#) - Playlist com diversos assuntos de Machine Learning. Recomenda-se os que tratam dos conteúdos do período.
- d. [Hands-on Machine Learning with Scikit-Learn, Keras, and TensorFlow](#) - Livro completo: Para o segundo período recomenda-se os capítulos 3, 4, 5, 6 e 7.
- e. [Handling imbalanced dataset in machine learning](#) - Vídeo sobre Undersampling e Oversampling.
- f. [Ensemble Learning](#) - Vídeo rápido que explica a ideia geral e dois dos métodos.
- g. [Ensemble methods: bagging, boosting and stacking](#) - Texto mais explicativo sobre Ensemble Learning.

### 3. Tarefas

- a. Base de Dados do Período - [Qualidade de Vinho](#)
- b. Atividade obrigatória: implementar um KNN utilizando somente numpy;
- c. Atividades sugeridas:
  - i. Classificação binária para prever se o vinho é tinto ou não;
  - ii. Classificação multiclasse para prever a qualidade do vinho;



- iii. Feature Importance - Aplicar ao algoritmo Random Forest e determinar quais features do dataset são mais importantes para o problema;
  - iv. Aplicar a normalização nos dados e verificar os efeitos nos modelos;
  - v. Aplicar as técnicas de undersampling e oversampling nos dados. Verificar e explicar os efeitos;
  - vi. Aplicar um dos métodos de Ensemble Learning e comparar os resultados.
- d. No meio do período, haverá uma reunião com o monitor para consolidação das informações;
- e. A entrega é individual e deverá ser colocada no seu GitHub pessoal até dia 28/07 às 23:59.