Guia docent [Codi UD] – [Sigles UD] – [Data Analysis and Machine Learning]

Unitat responsable:	Escola d'Enginyeria de Barcelona Est			
Unitat que imparteix:	Department of Mathematics			
Curs	2024	**Crèdits**	6	
Idiomes	Catalan			

PROFESSORAT	
Professorat responsable:	Francesc Pozo Montero
Altres:	

CAPACITATS PRÈVIES

To follow this course, students should have prior knowledge of **mathematics and statistics** (linear algebra, probability, regression), some experience in **Python programming** (libraries such as NumPy, Pandas, Matplotlib, Scikit-learn), and familiarity with **biomedical data analysis** (medical images, physiological signals, clinical data). Basic knowledge of **machine learning** (classification, regression, PCA) and strong analytical skills for interpreting results are recommended.

METODOLOGIES DOCENTS

The course uses theoretical content exposition (AF.1) and the resolution of exercises, problems, and cases (AF.2) for 20%, laboratory or simulation sessions (AF.7) for 20%, and individual and cooperative work for 60%.

OBJECTIUS D'APRENENTATGE DE L'ASSIGNATURA

Upon completing the **Data Analysis and Machine Learning** course, students will have acquired the necessary knowledge and skills to:

1. **Understand the fundamentals of machine learning applied to biomedicine**, including types of biomedical data, analysis techniques, and major supervised and unsupervised learning algorithms.
2. **Apply preprocessing and exploratory data analysis techniques to biomedical data**, such as normalization, feature selection, and dimensionality reduction, using tools like Scikit-learn and PyTorch.
3. **Develop and evaluate machine learning models for classification and regression** in biomedical problems, implementing logistic regression, decision trees, SVM, and neural networks.
4. **Implement deep learning models** for biomedical applications, including convolutional neural networks (CNNs) for medical image processing and recurrent neural networks (RNNs) for data sequences.
5. **Interpret the results obtained from artificial intelligence** models and assess their reliability using appropriate metrics (accuracy, sensitivity, AUC-ROC, etc.), cross-validation, and overfitting detection techniques.
6. **Apply interpretability and explainability methodologies** to ensure the ethical and safe use of artificial intelligence in biomedicine, using tools such as SHAP and LIME.
7. **Develop and deploy models in real biomedical environments**, understanding the challenges and practical considerations of transitioning a model from the experimentation phase to clinical application implementation.
8. **Integrate the acquired knowledge into a final project** where students will design and implement a machine learning solution applied to a specific biomedical problem.

HORES TOTALS DE DEDICACIÓ DE L'ESTUDIANTAT

Tipus	Hores	Percentatge
Hores activitats dirigides	6	4 %
Hores grup gran	30	20 %
Hores grup petit	24	16 %
Hores aprenentatge autònom	90	60 %
Dedicació total: 150 hores		

CONTINGUTS

Temari 1:	Fundamentals of Programming and Machine Learning

Descripció:

- Introduction to Python and essential libraries (NumPy, Pandas, Matplotlib, Scikit-learn, PyTorch).
- Basic concepts of machine learning: supervised vs. unsupervised learning, overfitting, cross-validation, evaluation metrics.

Activitats vinculades:
Practice: Exploratory analysis and preprocessing of biomedical data. A biomedical dataset (e.g., clinical records or physiological measurements) will be provided. Students will perform exploratory analysis, data cleaning, and normalization using Pandas, NumPy, and Scikit-learn.

Dedicació: 25 hores totals
Grup gran/Teoria: 5 h
Grup petit/Laboratori: 4 h
Activitats dirigides: 1 h
Aprenentatge autònom: 15 h

Temari 2:	Regression and Classification Models

Descripció:
- Linear and polynomial regression.
- Logistic regression and evaluation metrics in classification.
- Biomedical applications: prediction of clinical values, disease diagnosis.

Activitats vinculades:
Practice: Predicting clinical values with regression. Students will work with a patient dataset to predict a clinical variable (e.g., blood glucose levels). They will implement linear regression and Ridge/Lasso regression and compare performance using RMSE and R^2 metrics.

Dedicació: 25 hores totals
Grup gran/Teoria: 5 h
Grup petit/Laboratori: 4 h
Activitats dirigides: 1 h
Aprenentatge autònom: 15 h

Temari 3:	Dimensionality Reduction and Clustering

Descripció:
- PCA and advanced dimensionality reduction techniques.
- Clustering: K-means, DBSCAN, and hierarchical clustering.
- Interpretability and applications in biomedical data.

Activitats vinculades:
Practice: Visualization of biomedical data in 2D using PCA and t-SNE. A dataset with many variables (e.g., images transformed into vectors or genomic data) will be used to reduce dimensionality with PCA and t-SNE and visualize latent groups.

Dedicació: 20 hores totals
Grup gran/Teoria: 4 h
Grup petit/Laboratori: 3 h
Activitats dirigides: 1 h
Aprenentatge autònom: 12 h

Temari 4:	Advanced Models: Decision Trees and Ensemble Methods

Descripció:
- Decision trees, Random Forest, and Gradient Boosting.
- Feature selection and extraction.
- Applications in biomarkers and assisted diagnosis.

Activitats vinculades:
Problem-Solving: Model comparison for clinical anomaly detection. Various models (Decision Tree, Random Forest, Gradient Boosting) will be provided, and students will analyze which one performs best in terms of accuracy, recall, and feature importance in an anomaly detection case using ECGs.

Dedicació: 20 hores totals
Grup gran/Teoria: 4 h
Grup petit/Laboratori: 3 h
Activitats dirigides: 1 h
Aprenentatge autònom: 12 h

Temari 5:	Neural Networks and Deep Learning

Descripció:
- Introduction to deep neural networks.
- Fully Connected Neural Networks.
- Regularization and optimization in neural networks.

Activitats vinculades:
Practice: Building a neural network for medical diagnosis. PyTorch or TensorFlow will be used to construct an MLP for medical data classification (e.g., diabetes detection with the Pima Indians Diabetes dataset). Students will experiment with optimization, dropout, and L2 regularization.

Dedicació: 30 hores totals
Grup gran/Teoria: 6 h
Grup petit/Laboratori: 5 h
Activitats dirigides: 1 h
Aprenentatge autònom: 18 h

Temari 6:	Convolutional Networks and Applications in Biomedical Imaging

Descripció:
- Fundamentals of CNNs: convolutions, pooling, and classic architectures.

- Biomedical applications: medical image classification, segmentation, automated diagnosis.
- Model evaluation and ethical considerations.

Activitats vinculades:

Practice: Medical image classification using CNNs. A dataset of medical images (e.g., **chest X-rays for pneumonia detection**) will be used. Students will implement a basic CNN with **Keras/PyTorch**, analyze **confusion matrices**, **interpretability (Grad-CAM), and ethical biases**.

Dedicació: 30 hores totals

Grup gran/Teoria: 6 h
Grup petit/Laboratori: 5 h
Activitats dirigides: 1 h
Aprenentatge autònom: 18 h

SISTEMA DE QUALIFICACIÓ

Exemple:
First midterm exam = 35%
Second midterm exam = 45%
Projects, assignments, directed activities = 20%

Especificació

The evaluation will be conducted through faculty assessment. Students can pass the course through continuous assessment, consisting of two midterm exams (the first midterm midway through the course and the second during the official exam period) and the completion of projects, assignments, and directed activities. Each exam consists of a theoretical block (questions) and a practical block (problems), with a total duration of two hours.

There will be no re-assessment exam for this course.

BIBLIOGRAFIA

Bàsica:

Deprez, M., & Robinson, E. C. (2023). *Machine Learning for Biomedical Applications: With Scikit-Learn and PyTorch*. Academic Press.

Complementaria:

Kose, U., Deperlioglu, O., & Hemanth, D. J. (Eds.). (2023). Deep Learning for Biomedical Applications (1st ed.). Routledge.
Shaikh, T.A., Hakak, S., Rasool, T., & Wasid, M. (Eds.). (2023). Machine Learning and Artificial Intelligence in Healthcare Systems: Tools and Techniques (1st ed.). CRC Press.
Balas, V. E., et al. (Eds.). (2020). Handbook of Deep Learning in Biomedical Engineering (1st ed.). Academic Press.
Mohanty, S. N., Nalinipriya, G., Jena, O. P., & Sarkar, A. (2021). Machine Learning for Healthcare Applications. Wiley.
Kumar Rana, A., Sharma, S., Rana, S., & Chaudhary, R. (Eds.). (2023). Evolution of Machine Learning and Internet of Things Applications in Biomedical Engineering. Routledge.
Chandran, C. K. (Ed.). (2023). Machine Learning for Healthcare Systems. River Publishers[9].
Singh, S. K., & Sinha, P. K. (Eds.). (2021). Machine Learning in Healthcare: Fundamentals and Recent Applications. CRC Press[12].

RECURSOS

Altres recursos:

Exemple:
Class material available on ATENEA.
https://ocw.mit.edu/courses/6-867-machine-learning-fall-2006/