



### **Machine Learning (MINT)**

Guía docente 2023-24

## **PRESENTACIÓN**

### Breve descripción:

The recent increase in computing power, together with new parallel computing techniques, have opened up a huge range of potential applications in multiple industries.

This subject will study, from a theoretical-practical point of view, the newest technologies of both machine and deep learning (ML and DL) research for the search for applications in different fields.

- **Titulación:** Máster en Innovación Tecnológica
- **Módulo/Materia:** Tecnologías Emergentes (MINT) / Fundamentos avanzados de análisis de datos y aprendizaje automático
- **ECTS:** 5 ECTS
- **Curso, semestre:** 1º, Primero
- **Carácter:** Obligatorio
- **Profesorado:**
  - [Rubio Díaz-Cordovés, Ángel](#) - Email: [arubio@tecnun.es](mailto:arubio@tecnun.es) / Catedrático
  - Díez Parra, Mikel - Email: [mdiezparra@external.unav.es](mailto:mdiezparra@external.unav.es) / Invitado (Colab. Docente)
  - [Granada Echeverria, Imanol](#) - Email: [igranada@external.unav.es](mailto:igranada@external.unav.es) / Invitado (Colab. Docente)
  - [Rodríguez Ferradas, María Isabel](#) - Email: [mirodriguez@tecnun.es](mailto:mirodriguez@tecnun.es) / Profesor contratado doctor
- **Idioma:** Inglés

## **COMPETENCIAS**

- CG1 - Integrar visión estratégica y tecnología para generar nuevos modelos de negocio.
- CG2 - Coordinar grupos de trabajo multidisciplinares para desarrollar procesos de transformación basados en tecnologías emergentes.
- CB6 - Poseer y comprender conocimientos que aporten una base u oportunidad de ser originales en el desarrollo y/o aplicación de ideas, a menudo en un contexto de investigación
- CB7 - Que los estudiantes sepan aplicar los conocimientos adquiridos y su capacidad de resolución de problemas en entornos nuevos o poco conocidos dentro de contextos más amplios (o multidisciplinares) relacionados con su área de estudio
- CB9 - Que los estudiantes sepan comunicar sus conclusiones y los conocimientos y razones últimas que las sustentan a públicos especializados y no especializados de un modo claro y sin ambigüedades
- CB10 - Que los estudiantes posean las habilidades de aprendizaje que les permitan continuar estudiando de un modo que habrá de ser en gran medida autodirigido o autónomo.
- CE6 - Conocer y aplicar las técnicas de análisis de datos y los principales modelos de aprendizaje automático en diferentes casos de uso.



# Universidad de Navarra

- CE7 - Conocer las principales herramientas, tecnologías y entornos de desarrollo disponibles para implementar soluciones de inteligencia artificial.
- CE8 - Evaluar ventajas e inconvenientes de sistemas de captación, depuración, visualización de datos.

## PROGRAMA

Machine Learning Course Syllabus:

Machine Learning Overview:

- Introduction to supervised and unsupervised learning
- Train, Validation, Test Paradigm
- Understanding the importance of data splitting
- Evaluating model performance

Introduction to R

- Overview of the R programming language
- Basic data manipulation and analysis in R
- Data visualization using R

Introduction to Python

- Overview of Python programming language
- Basics of Python for machine learning

Supervised Learning: Regression

- Understanding regression models
- Building and evaluating regression models

Supervised Learning: Classification

- Logistic Regression, SVM, and Multi-class Classification
  - Logistic regression for binary classification
  - Support Vector Machines (SVM) for binary and multi-class classification
- Classification Trees, Random Forests, and XGBoost
  - Decision tree-based algorithms for classification
  - Ensemble methods and boosting techniques

Unsupervised learning

- Introduction to unsupervised learning algorithms
- Clustering techniques for identifying patterns in data
- Dimensionality reduction techniques
- Introduction to recommender systems

Case Studies and Overarching Project

- Applying machine learning techniques to real-world scenarios

Natural Language Processing (NLP) Workshop

- Analyzing and processing textual data



# Universidad de Navarra

- Computer Vision (CV) Workshop
  - Applying machine learning to visual data analysis
- Workshop on Large Language Models (chatGPT like)

## ACTIVIDADES FORMATIVAS

### EVALUACIÓN

#### CONVOCATORIA ORDINARIA

60% Proyecto Paraguas.

The project will include at least one machine learning problem. The student can select a dataset (see the end of this document that includes several repositories) and propose a data-driven task that could be solved using any machine learning technique, even if it was not seen in class.

The deliverables will be:

Milestones:

H1: Online discussion with Angel Rubio. (optional)

Sprint 1: Selection of a dataset. Preliminary ideas about the data-driven business concept. Discussion on the ML technique to be used.

Deliverable: Brief report explaining the aforementioned information.

H2: Revision of the project with the lecturer. Angel Rubio. 10% (compulsory)

More in-depth discussion of the preliminary ideas.

Deliverable: Incremental report enriched by the discussion.

H3: Online discussion with Imanol Granada. 25%. (compulsory)

Sprint 2: Exploratory analysis of the dataset. Definition of the target and features. Data-driven hypotheses and (optional) preliminary predictive baseline.

Deliverable: Incremental report that will consist mostly of displaying properly the data to be analyzed.

H4: Angel Rubio. 25% (compulsory)

Sprint 3: Final technical report (pdf) and all source code (R or Python) related to the ML or DL problem.

Deliverable: final report (50%)

This final report should include:

- Source code used with a README file explaining each code file.
- A technical report (PDF) will include at least:



1. Business problem definition.
2. Description of the dataset used, features, exploratory analysis, evaluation strategy and hypotheses.
3. Description of the steps of the ML analysis, explain data-based insights and results.
4. Conclusions, limitations and future lines.

The final proposal should take into account what drives added value to customers, as well as ensure the viability of the project. The final proposal will be graded considering:

- Technical feasibility of the project.
- Innovativeness of the proposal. Profit for the users.
- Appropriateness of the employed ML algorithm or technique.
- Well-written documentation.
- Source code should be clear, ordered and commented\*.
- Analysis of limitations.

\* Any piece of code published on the internet can be used with two conditions:

- (1) code MUST be referenced both in the documentation and in the script;
- (2) every piece of code MUST be understood by the students.

## 40% Otras entregas y participación.

- Despues de cada tema se pedirá hacer un pequeño test para asentar los conocimientos.

## CONVOCATORIA EXTRAORDINARIA

- Machine learning project (different form the previous one) (60%)
- Exam on Machine Learning (40%).

## HORARIOS DE ATENCIÓN

Dr Angel Rubio ([arubio@unav.es](mailto:arubio@unav.es))

- Horario de tutoria: Concertar teleconferencia con el profesor.

## BIBLIOGRAFÍA

### Annex: (possible sources of data)

General sources:

<https://github.com/awesomedata/awesome-public-datasets>

<https://www.kaggle.com/datasets> (use the opendatasets library for downloading datasets)

<https://course.fast.ai/datasets>

<https://www.kaggle.com/competitions> (check the "Completed" tab)



Universidad  
de Navarra

<https://archive.ics.uci.edu/ml/index.php>

<https://datasetsearch.research.google.com/>

World Covid-19 Dataset

<https://www.kaggle.com/imdevskp/corona-virus-report>

AI for Social Good - Google AI

<https://ai.google/social-good/>