

GUIA DOCENTE
MUFIBAM-MUFIBAR

Curso 2024-2025

ASIGNATURA

- **Nombre:** Ciencia de datos
- **Tipo de asignatura:** y tú de asignatura
- **Créditos:** 4 ECTS
- **Idioma de docencia:** CASTELLANO
- **Coordinador de la asignatura:** LUZ MARY PINZON
- **Profesor de la asignatura:** LUZ MARY PINZON y FERRAN CARRASCOSA

PRESENTACIÓN DE LA ASIGNATURA

Este curso es una introducción al aprendizaje automático, con aplicaciones en finanzas. Primero nos enfocamos en problemas que son independientes del contexto o área de aplicación específicos; por ejemplo: cómo determinar qué modelo funciona mejor entre un conjunto de candidatos, cómo estimar correctamente la precisión de un modelo, cómo hacer que nuestros modelos se generalicen bien a nuevos datos.

Presentamos un marco estadístico mínimo utilizado para comprender el aprendizaje y luego profundizamos en problemas principales del aprendizaje supervisado

OBJETIVOS

- **Competencias generales:**
 - Analizar y manipular datos: lectura, limpieza, visualización y exploración de datos.
 - Aprender desde una perspectiva matemática y estadística; Problemas de aprendizaje supervisado, no supervisado.
 - Modelado: selección, evaluación y comparación de modelos.
 - Problemas de regresión y clasificación desde el punto de vista del Machine Learning.
 - Machine Learning con pipeline usando Python.
- **Métodos y conceptos específicos:**
 - Estimación de un modelos, pérdidas, sesgo y varianza, sobreajuste y desajuste.
 - Entrenamiento de un modelo: optimización
 - Realización de la selección del modelo mediante validación cruzada
- **Herramientas:**
 - Usar Python and Jupyter notebooks
 - Familiaridad con librerías, pandas, numpy, pyplot, sklearn.

Prerequisitos:

Haber superado el curso introductorio de Python(Brush up) y Datos masivos

Metodología docente

La metodología utilizada será híbrida. Las clases son tanto conferencias por parte del profesor como laboratorios explicados a través de ejercicios y casos prácticos, Los laboratorios complementan la teoría y familiarizan a los estudiantes con el lenguaje de programación Python y sus principales bibliotecas de ciencia de datos.

Las competencias, los resultados del aprendizaje, los elementos de evaluación y la calidad del proceso de aprendizaje incluidos en este Plan Docente no se verán afectados si durante el trimestre académico el modelo docente tiene que cambiar a un modelo híbrido (combinación de presencial y sesiones en línea) o a un modelo completo en línea.

Evaluación del curso

Evaluación continua: ● Quices. 10% de la nota.
● Laboratorios y talleres. 40% de la nota.

Examen final. 50% de la nota.

La nota del examen final debe ser mayor o igual a 4 para que haga computo con la evaluación continua, de lo contrario la nota del examen será la nota final.

Solo el examen final se transfiere a una eventual recuperación. No hay recuperación para la evaluación continua (los quices o proyectos que conforman la evaluación a lo largo de las clases.).

Para aprobar la asignatura se necesitan mínimo 50 puntos sobre 100.

A los estudiantes que no superen la evaluación del curso se les permite hacer un segundo examen ó examen de recuperación (retake) cuya nota reemplaza la nota del examen.

En caso de inconveniente para presentar el examen, deberá informar al profesor correspondiente y al (los) director (es) del programa para que estudien la posibilidad de reprogramar el examen.

Se requiere que los estudiantes asistan al 80% de las clases.

PROFESORADO

Luz Mary Pinzón S.

Doctorado en Estadística, Universidad de Salamanca-España. Magíster en Estadística, Universidad Nacional de Colombia. Licenciada en Matemáticas, Universidad Pedagógica Nacional de Colombia.

Experta en modelos estadísticos aplicados, análisis e inferencia multivariada.

Aplicaciones de Machine Learning en Python y R.

Consultora Senior en Estadística aplicada con experiencia en medio ambiente, educación, investigación social y marketing.

Profesora de estadística y matemáticas aplicadas a la ingeniería, economía, contaduría, marketing, finanzas.

Ferran Carrascosa Mallafrè

Profesor Universidad Pompeu Fabra.

Gerente Data Scientist del equipo de Risk Analytics en CaixaBank. Apasionado del análisis de datos con veinte años de experiencia laboral, inicialmente 5 años en Estudios de Mercado y posteriormente 10 años en la consultora AIS de Aplicaciones de Inteligencia Artificial en los sectores de Banca y Retail. Actualmente como gerente metodológico en el equipo de innovación en Big Data en Banca. Quince años de experiencia como Profesor Asociado en el departamento de Economía y Empresa en la Universitat Pompeu Fabra.

BIBLIOGRAFIA (obligatoria/ recomendada)

1. John Hull. Machine Learning in Business. Second edition, 2020.
2. <https://ipfs.io/ipfs/bafykbzaceaeemelf2lv3wy7636ctftoenxl4aevykdj43msllnrm32olcss?filename=Jake%20VanderPlas%20-%20Python%20Data%20Science%20Handbook%20Essential%20Tools%20for%20Working%20with%20Data-O%27Reilly%20Media%20%282016%29.pdf>
3. Brown, Eric D., 2014. Analysis of Twitter Messages for Sentiment. Masters Theses & Doctoral Dissertations. 291.
4. R. Mitchell. Web Scraping with Python. O'Reilly Media, Inc. 2nd ed.; 2018. Disponible el código en: <https://github.com/REMitchell/python-scraping>