#### Asignatura: INGENIERÍA DE DATOS

**DATOS PRELIMINARES**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Temas** | **H.** | **Conf.** | **C.P.** | **Sem.** | **Lab.** | **eval.** |
| Metodologías para la ciencia de Datos | 4 | 2 |  | 2 |  |  |
| El ecosistema de la ingeniería de datos | 4 | 2 |  | 2 |  |  |
| Tecnologías para los flujos de datos | 18 | 8 | 6 |  | 4 |  |
| Integración de sistemas | 18 | 8 | 6 |  | 4 |  |
| Devops | 14 | 6 | 4 |  | 4 |  |
| Optimización y detección de errores | 6 | 2 | 2 |  | 2 |  |
| **Totales** | **64** | **28** | **18** | **4** | **14** |  |

**FUNDAMENTACIÓN DE LA ASIGNATURA**

La ingeniería de datos se encarga de organizar, tanto metodológica como tecnológicamente, los procesos de ciencia de datos. De esta manera, en concordancia con las metodologías existentes, es responsabilidad de la ingeniería de datos organizar un flujo de datos estándar para su procesamiento y análisis. Este flujo se debe entender como la ingestión de datos, su almacenamiento, su preprocesamiento, su transformación, procesamiento y presentación.

Por lo tanto, la ingeniería de datos dota a los estudiantes de una comprensión completa de las metodologías, las tecnologías y las herramientas más eficientes, para la captura y el procesamiento de conjuntos de datos complejos con confiabilidad. Además, proporciona conocimientos para la implantación y mantenimiento de estos flujos de datos que combinan múltiples metodologías, datos y tecnologías.

**OBJETIVOS GENERALES DE LA ASIGNATURA**

1. Caracterizar las diferentes metodologías que sustentan el desarrollo de proyectos de ingeniería de datos.
2. Valorar las posibilidades, ventajas y desventajas de las distintas metodologías, herramientas y tecnologías que intervienen en el ciclo de vida del dato y su análisis.
3. Identificar y comprender los distintos roles que existen en los procesos de

ciencia de datos,

1. Diseñar tecnológicamente flujos ( pipelines ) de datos que pueden ir desde la ingestión de datos hasta la presentación de los resultados, pasando por el almacenamiento, pre-procesamiento, transformación y procesamiento.
2. Implantar, gestionar y administrar los sistemas que resultan a partir de diseño concretos de flujos de datos.
3. Optimizar el funcionamiento de los sistemas que lleven a la práctica los flujos de datos
4. Diagnosticar y solucionar los errores en los sistemas que lleven a la práctica los flujos de datos

**CONTENIDOS BÁSICOS DE LA ASIGNATURA**

**Conocimientos esenciales a adquirir**

El ecosistema de datos. Metodologías para la ciencia de datos. El ecosistema de la ingeniería de datos. Responsabilidades y habilidades dentro de la ciencia de datos. Tipos de datos, formatos y fuentes de datos. Repositorios de Datos.

Sistemas para el almacenamiento y gestión de datos (*databases, data*

*warehouses, data marts, data lakes*). ETL/ ELT y su relación con los flujos

(*pipelines*) de datos. Sistemas de integración de datos. Sistemas para el

procesamiento de datos. Sistemas y herramientas para la transformación y

consulta de datos. Modelos y herramientas para la interoperabilidad de sistemas.

Diseño y arquitectura de plataformas de datos. Sistemas en la nube. Devops.

Optimización y detección de errores. Acercamiento a los modos de integración de los datos. Problemas frecuentes en los escenarios de integración de información.

**Habilidades principales a dominar**

1. Caracterizar el ecosistema de la ingeniería de datos
2. Manipular los distintos formatos de datos
3. Acceder a fuentes heterogéneas de datos
4. Instalar, utilizar y gestionar sistemas que permitan el almacenamiento, la integración y el procesamiento de datos.
5. Utilizar lenguajes para la consulta de datos
6. Implantar y gestionar sistemas (*devops*) de manera local, en red o en la nube
7. Diseñar y realizar experimentos para evaluar rendimiento así como seleccionar métricas de rendimiento
8. Participar en equipos para la definición, creación e implantación de productos de ciencia de datos
9. Redactar documentos, exponer los resultados y participar en foros de discusión. Así como leer e interpretar bibliografía en idioma inglés.

**Valores fundamentales de la carrera a los que tributa**

La asignatura contribuye mediante su sistema de trabajo al fomento de los valores de responsabilidad ante las tareas asignadas; de honestidad, resaltando el sentido de pertenencia y deber con la sociedad; dignidad, mediante el compromiso revolucionario en el cumplimiento de las funciones como profesional.

Desarrollar las cualidades de tenacidad y dedicación necesarias en el trabajo de ciencia de datos.

Acometer la solución de problemas con una amplia perspectiva científica, basada en el razonamiento lógico y el pensamiento abstracto y un alto rigor en la obtención de resultados, honestidad, responsabilidad, solidaridad, basado en la formación humanista y la ética de la profesión.

La motivación, la creatividad y el amor hacia la profesión y el compromiso con la estética de los resultados.

Trabajar de manera independiente y colaborativa, con constancia y organización en el estudio.

El desarrollo de la responsabilidad social y del espíritu crítico a través de la naturaleza conflictiva de los modelos reales que se estudian en las asignaturas.

Garantizar la protección de los datos, la veracidad de la información y la privacidad de quienes la proveen.

**INDICACIONES METODOLÓGICAS Y DE ORGANIZACIÓN DE LA ASIGNATURA**

En esta asignatura se garantiza el establecimiento de la infraestructura

computacional básica para desarrollar el ciclo completo de la ciencia de datos. Es por ello que se debe impartir posteriormente a las asignaturas relacionadas con los sistemas computacionales, las redes de computadoras y el procesamiento de grandes volúmenes de datos, así como aquellas asignaturas que, desde el punto de vista teórico marcan las pautas de los procesos de ciencia de datos. Su evaluación ha de proponerse en el ámbito de desarrollar la infraestructura para la solución de un problema de ciencia de datos real.

Para el desarrollo de esta asignatura se proponen los siguientes tipos de

actividades docentes:

* **Clases de introducción de nuevos contenidos.** (Conferencias). Se explicará la esencia de los métodos que resuelven los problemas que son objeto de estudio, y se enfatizará en las particularidades de los algoritmos que se elijan para ilustrar cada tipo.
* **Clases de ejercitación.** (Clases Prácticas). Se centrarán, fundamentalmente, en la solución de problemas típicos de la asignatura, mediante la discusión de ejercicios propuestos, de forma individual o en equipos. Enfrentarse a problemas prácticos para que los estudiantes aprendan a reconocer dónde utilizar los conocimientos adquiridos en la asignatura.
* **Seminarios.** Los estudiantes expondrán temas orientados como estudio independiente o resultados de investigación orientadas. Se propone que los temas de los seminarios estén relacionados con aplicaciones de los contenidos de la asignatura a la Ciencia de Datos.
* **Laboratorios.** (Clases prácticas). Se centrarán fundamentalmente en la utilización de los recursos computacionales para la solución de problemas, ya sea desde la implementación de los métodos y algoritmos estudiados en conferencia, hasta su aplicación a problemas prácticos a partir de su ejecución computacional.

Se recomienda la utilización de materiales digitales para la bibliografía con el

objetivo de mantener actualizados a los estudiantes y la utilización de materiales en inglés para fomentar el aprendizaje y la preparación de los estudiantes en este idioma.

Como complemento de esta asignatura se propone ofertar cursos optativos

relacionados con tecnologías y herramientas concretas para el despliegue de

sistemas de análisis de datos en la nube, tales como Amazon AWS, Google Cloud Engine, Microsoft Azure, entre otros.

Con vistas a fomentar la interdisciplinaridad, se recomienda valorar la realización de evaluaciones integradas entre la presente asignatura, y otras asignaturas de la disciplina Programación y Sistemas de Información que traten los temas relacionados con el desarrollo de sistemas de recuperación de información, como un dominio concreto donde aplicar los conocimientos y metodologías de la ingeniería de datos.

**Sistema de evaluación**

Para la evaluación de esta asignatura se proponen los siguientes tipos de actividades evaluativas.

* **Evaluaciones parciales** en donde se demuestren las habilidades que han alcanzado los estudiantes, con una periodicidad que determine el colectivo de la asignatura. Las evaluaciones sistemáticas evalúan conocimientos y habilidades adquiridas por los estudiantes. Las mismas pueden ser aplicadas a través de preguntas escritas, orales, reportes de lectura u otras.
* **Tarea extraclase** de mayor alcance a evaluar y defender frente al profesor. El desarrollo de este trabajo puede unificarse con las evaluaciones sistemáticas del trabajo de los estudiantes de forma que puedan utilizar las evaluaciones sistemáticas para corregir el trabajo.

**BIBLIOGRAFÍA**

**BÁSICA**

1. Joe Reis, Matt Housley. Fundamentals of Data Engineering. O’reilly 2022

2. Jesse Anderson. Data Teams: A Unified Management Model for Successful Data-Focused Teams. Apress, 2020

**COMPLEMENTARIA**

3. Roger Lee. Big Data, Cloud Computing, and Data Science Engineering. Studies in Computational Intelligence, 2020