

Facultad de Ingeniería / Ingeniería Civil Plan Común

Nombre de la actividad curricular: Fundamentos de Ciencia de Datos
 Código:
 Créditos: 10 UAH / 6 SCT
 Carácter: Obligatorio
 Prerrequisitos: Probabilidades y Estadística
 Tipo: Curso / Laboratorio
 Horas cronológicas de dedicación Docencia directa: 3 Trabajo autónomo: 7

I. DESCRIPCIÓN

Esta actividad curricular, pertenece área de **formación disciplinar y/o profesional (formación e iniciación a la disciplina y/o profesión)**, y se encuentra ubicada en el **ciclo básico** del plan de estudio. Busca desarrollar en el estudiantado capacidades para buscar soluciones a problemas y asistir la toma de decisiones a través de las metodologías básicas basadas en el uso de datos. Esto implica conocer la terminología básica, los conceptos técnicos, los desafíos tecnológicos de implementación y las implicancias éticas del uso de datos. Esta actividad curricular promueve un aprendizaje profundo de las estrategias fundamentales de la ciencia de datos sobre la base del autoaprendizaje, la autoevaluación, la autonomía y la autorregulación, dejando en segundo plano las herramientas; todo esto debido a la importancia de las primeras y la naturaleza cambiante de las últimas.

La presente actividad curricular contribuye al logro de los siguientes desempeños del perfil de egreso:

Ingeniería Civil Industrial	Ingeniería Civil en Informática
1. Concibe soluciones a problemas complejos de ingeniería aplicando principios de matemáticas, ciencias e ingeniería, investigación de operaciones, sistemas productivos, estadística, ingeniería económica y factores humanos a fin de enfrentar los desafíos emergentes de la sociedad en los ámbitos económicos, ambientales y sociales.	1. Concibe soluciones a problemas complejos de ingeniería aplicando principios de computación, matemáticas, ciencias e ingeniería, ciencias de la computación, ingeniería de software e interfaz hombre-máquina a fin de enfrentar los desafíos emergentes de la sociedad en los ámbitos económicos, tecnológicos, ambientales y sociales.
2. Diseña sistemas integrados de producción, operación, y servicios considerando la salud pública, la seguridad, y el bienestar, así como factores globales, culturales, sociales, ambientales y económicos, para facilitar la toma de decisiones, mejorando la competitividad de las organizaciones y la calidad de vida de las personas.	2. Diseña sistemas tecnológicos considerando la salud pública, la seguridad, y el bienestar, así como factores globales, culturales, sociales, ambientales y económicos, para facilitar la toma de decisiones, mejorando la competitividad de las organizaciones y la calidad de vida de las personas.

Facultad de Ingeniería / Ingeniería Civil Plan Común

II. PROPÓSITOS DE APRENDIZAJE

Al término de esta actividad curricular, el estudiantado será capaz de:

1. Analizar de forma crítica cuándo y cómo utilizar soluciones basadas en datos en problemas reales, considerando las oportunidades y dilemas éticos del uso de estas técnicas.
2. Implementar soluciones efectivas de principio a fin a problemas utilizando las estrategias más básicas basadas en datos.
3. Juzgar de forma crítica las limitaciones y desafíos pendientes del uso de este tipo de metodologías basadas en datos.
4. Construir plataformas para interactuar, refinar y visualizar los resultados de los métodos básicos basados en la ciencia de datos.

III. CONTENIDOS

Unidad I: Ciencia de Datos y Big Data

1. Qué es Big Data: definiciones, desafíos e infraestructura
2. Qué es ciencia de datos: visión de la disciplina, tareas, labores y toma de decisiones
3. Ingeniería de datos, minería de datos, analista de datos, ciencia de datos
4. Modelos predictivos, discriminativos y generativos
5. Modelos supervisados y no-supervisados
6. Repaso de distribuciones, media y varianza muestral, normalidad
7. Principios básicos de estadística bayesiana

Unidad II: Temas éticos en la ciencia de datos

1. Sesgos y diversidad
2. Transparencia e interpretabilidad: hacia algoritmos justos
3. Privacidad y protección: límites de la anonimización

Unidad III: Estadística descriptiva

1. Actividad: desafío de reporte y descripción de datos reales
2. Tipos, formatos, fuentes y métodos de extracción de datos
3. Exploración inicial de datos
4. Preparación de los datos – tratamiento, limpieza e integración de datos de distintas fuentes
5. Construcción de nuevas variables e índices

Facultad de Ingeniería / Ingeniería Civil Plan Común

Unidad IV: Modelos predictivos – Regresión

1. Actividad: desafío de predicción con alguna base de datos pública
2. Mínimos cuadrados
3. Formulación bayesiana conexión con problemas inversos
4. Interpretación de los resultados

Unidad V: Espacios de características

1. Generación de características de los datos y concatenación
2. Normalización
3. Selección de características
4. Reducción de dimensionalidad (PCA)

Unidad VI: Modelos discriminativos y generativos – Problema de clasificación

1. Actividad: desafío de clasificación
2. Modelos supervisados
3. Modelos no-supervisados
4. Métricas de rendimiento – Curvas ROC, AUOC, matriz de confusión
5. Balance de las clases y técnicas de mitigación de desbalance

Unidad VII: Visualización de Datos

1. Modelos y herramientas para visualización de datos de distinta naturaleza
2. Diseño y construcción de interfaces de reporte de indicadores de desempeño
3. Desafío final: construcción sistema end-to-end

IV. METODOLOGÍA

- Clases expositivas y participativas.
- Aprendizaje entre pares
- Aprendizaje basado en problemas.
- Aprendizaje basado en proyectos

V. EVALUACIÓN DE APRENDIZAJES

- Informes escritos
- Proyecto aplicado/diseñado
- Presentaciones orales

Facultad de Ingeniería / Ingeniería Civil Plan Común

VI. RECURSOS PEDAGÓGICOS

Recursos bibliográficos

Bibliografía básica

1. Joel Grus. (2019). Data Science from Scratch: First Principles with Python. Sebastopol, CA, Estado Unidos, Editorial O'Reilly.
2. Wes McKinney (2018). Python for Data Analysis: Data wrangling with Pandas, Numpy and IPython. Sebastopol, CA, Estados Unidos, Editorial O'Reilly.
3. Bobadilla, J. (2020). Machine Learning y Deep Learning: Usando Python, Scikit y Keras. 1. Paracuellos de Jarama, Madrid, RA-MA Editorial.

Bibliografía complementaria

1. Menoyo Ros, D. García López, E. y García Cabot, A. (2021). Fundamentos de la ciencia de datos. Alcalá de Henares, España, Editorial Universidad de Alcalá.
2. Vicente Vírseda, J. A. González Arias, J. y Parra Rodríguez, F. J. (2019). Métodos de Data Science aplicados a la Economía y a la Dirección y Administración de Empresas. Madrid, UNED - Universidad Nacional de Educación a Distancia.