



UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO
FACULTAD DE INGENIERÍA



TEMAS SELECTOS DE
INGENIERÍA EN COMPUTACIÓN

II

CIENCIA DE DATOS PARA LA
TOMA DE DECISIONES EN LAS
ORGANIZACIONES

Asignatura

2957

Clave

10

Semestre

8

Creditos

INGENIERÍA ELÉCTRICA

División

COMPUTACIÓN

Departamento

INGENIERÍA EN COMPUTACIÓN

Carrera(s) en que se imparte

Asignatura:

Obligatoria

Optativa

Horas/semana:

Teóricas

Prácticas

Total

Horas/semestre:

Teóricas

Prácticas

Total

Modalidad: Curso teórico

Seriación obligatoria antecedente: Ninguna

Seriación obligatoria consecuente: Ninguna

Descripción del
curso:

Este es un curso introductorio a la Ciencia de Datos que se enfoca en el uso de herramientas y tecnologías de código abierto ampliamente utilizadas en esta industria, como Python y sus bibliotecas para el análisis de datos (pandas, numpy, matplotlib, seaborn, entre otras). También se utilizarán plataformas en línea como Google Colab para facilitar el trabajo con datos y la presentación de resultados a la organización, gestión de expectativas y stakeholders, con un enfoque de administración de proyectos de ciencia de datos. Finalmente se realizará un proyecto que integre los conocimientos adquiridos durante todo el curso.

Objetivo(s) del curso:

Este curso tiene como objetivo proporcionar una introducción a la Ciencia de Datos, que es una materia de carácter interdisciplinario que combina el uso de la estadística, matemáticas, programación y conocimientos de dominio/negocio para extraer información valiosa a partir de datos. Durante el curso, los estudiantes aprenderán sobre los conceptos, técnicas y herramientas fundamentales e introductorios de la Ciencia de Datos, incluyendo: limpieza, preparación, exploración, visualización y modelado estadístico de datos.

Temario

NÚM.	NOMBRE	HORAS
1	Introducción a la Ciencia de Datos y su Aplicación en Toma de Decisiones	4.0
2	Fundamentos de Programación en Python	4.0
3	Manipulación de Datos con Python - Parte 1	4.0
4	Manipulación de Datos con Python - Parte 2	4.0

5	Fundamentos de Estadística para Ciencia de Datos	4.0
6	Visualización de Datos	4.0
7	Exploración de Datos y Análisis Descriptivo	4.0
8	Introducción a Modelos de Machine Learning	4.0
9	Modelos de Clasificación: Conceptos y Ejemplos	4.0
10	Estrategias Avanzadas de Limpieza y Transformación de Datos	4.0
11	Preparación para el Proyecto Final y Trabajo en Equipo	4.0
12	Desarrollo de un Pipeline de Análisis de Datos	4.0
13	Introducción a la Toma de Decisiones Basada en Datos	4.0
14	Proyecto Final - Desarrollo y Asesoría	8.0
	Actividades prácticas	0.0
	Total	64.0

TEMARIO DESGLOSADO

1. Tema 1: Introducción a la Ciencia de Datos y su Aplicación en Toma de Decisiones
Objetivo: Comprender el concepto y ciclo de vida de la Ciencia de Datos, explorando su aplicación en diversos sectores, así como familiarizarse con las herramientas principales del curso.
Contenido:
 1. ¿Qué es la Ciencia de Datos? Ciclo de vida de un proyecto.
 2. Aplicación de la Ciencia de Datos en diferentes sectores: social, empresarial.
 3. Introducción a las herramientas del curso: Python, Notebooks (Libros de trabajo)
 4. Soluciones que escalan ¿Por qué Google Colab?

2. Tema 2: Fundamentos de Programación en Python
Objetivo: Adquirir conocimientos básicos de Python, incluyendo variables, tipos de datos y estructuras de control, y establecer un entorno de trabajo efectivo con Google Colab.
Contenido:
 1. Introducción a Python: variables, tipos de datos, estructuras de control.
 2. Entorno de trabajo: Google Colab, Jupyter Notebooks.
 3. Entornos de trabajo a gran escala: Google Vertex AI: Colab Enterprise o Workbench

3. Tema 3: Manipulación de Datos con Python - Parte 1
Objetivo: Desarrollar habilidades para trabajar con objetos de datos (CSV, Excel), Librería Pandas para la carga, manipulación y selección de datos en estructuras como Series y DataFrames.
Contenido:
 1. Uso básico de la librería Pandas: carga de datos, Series, DataFrames.
 2. Indexación y selección de datos.
 3. Entorno de trabajo: Google Colab, Jupyter Notebooks.

4. Tema 4: Manipulación de Datos con Python - Parte 2
Objetivo: Aplicar técnicas de limpieza y transformación de datos para preparar conjuntos de datos utilizables, resolviendo problemas comunes como valores nulos y duplicados, agregando nuevos objetos de datos (Parquet, Json).
Contenido:

1. Limpieza de datos: valores nulos, duplicados, cambio de tipos de datos.
 2. Transformación de datos: agregar, filtrar, modificar.
 3. Segmentación de datos para el negocio.
 4. Hallazgos básicos en los datos.
5. Tema 5: Fundamentos de Estadística para Ciencia de Datos
Objetivo: Comprender conceptos clave de estadística descriptiva y utilizar herramientas gráficas iniciales como histogramas y boxplots para resumir datos.
Contenido:
1. Estadísticas descriptivas: medidas de tendencia central y dispersión.
 2. Visualización inicial de datos: histogramas, boxplots.
 3. Generación de hallazgos básicos para negocio.
6. Tema 6: Visualización de Datos
Objetivo: Crear visualizaciones efectivas utilizando matplotlib y seaborn, seleccionando tipos de gráficos adecuados para los diferentes tipos de datos.
Contenido:
1. Creación de gráficos con matplotlib y seaborn: histogramas, scatter plots.
 2. Selección adecuada de gráficos según los datos.
7. Tema 7: Exploración de Datos y Análisis Descriptivo
Objetivo: Llevar a cabo un análisis exploratorio de datos (EDA) para identificar patrones, valores atípicos y relaciones dentro de un conjunto de datos.
Contenido:
1. Análisis exploratorio de datos (EDA).
 2. Identificación de patrones, outliers, y relaciones.
 3. Contar historias con Datos.
8. Tema 8: Introducción a Modelos de Aprendizaje Automático (ML)
Objetivo: Brindar un contexto sobre el Aprendizaje Automático y su aplicación en los negocios. Diferenciar entre aprendizaje supervisado y no supervisado, e introducir el uso de scikit-learn para entrenar y evaluar modelos básicos.
Contenido:
1. Contexto del Aprendizaje Automático.
 2. Aprendizaje Automático en los negocios.
 3. Diferencias entre aprendizaje supervisado y no supervisado.
 4. Introducción a scikit-learn para entrenar y evaluar modelos.
9. Tema 9: Modelos de Clasificación: Conceptos y Ejemplos
Objetivo: Explorar los fundamentos de los modelos de clasificación como regresión logística y árboles de decisión, además de comprender las métricas de evaluación como precisión, rerecuerdo (recall) y F1-score.
Contenido:
1. Introducción a los modelos de clasificación: regresión logística, árboles de decisión.
 2. Evaluación de modelos: precisión, recall, F1-score.
10. Tema 10: Estrategias Avanzadas de Limpieza y Transformación de Datos.
Objetivo: Aplicar técnicas avanzadas de limpieza y transformación de datos con Python, para el manejo de valores atípicos, normalización y escalado, para preparar conjuntos de datos para análisis y modelado.
Contenido:
1. ¿Por qué es tan valiosa la limpieza de datos?
 2. Limpieza avanzada: manejo de datos faltantes y atípicos, normalización y escalado.
 3. Preparación de datos para modelado.
11. Tema 11: Preparación para el Proyecto Final y Trabajo en Equipo
Objetivo: Diseñar y planificar un proyecto de Ciencia de Datos, definiendo objetivos claros, métricas de éxito y roles dentro del equipo de trabajo.
Contenido:
1. Planificación y estructuración de un proyecto de Ciencia de Datos.
 2. Revisión de los objetivos, métricas de éxito, y fuentes de datos.
 3. Trabajo en equipo: roles y colaboración.

12. Tema 12: Desarrollo de un Flujo de Trabajo (Pipeline) de Análisis de Datos
 Objetivo: Construir un pipeline funcional en Python que automatice procesos de limpieza, análisis y visualización de datos.
 Contenido:
 1. Conceptos básicos de DataOps
 2. Creación de un pipeline para análisis de datos en Python.
 3. Automatización de tareas de limpieza, análisis y visualización.
13. Tema 13: Introducción a la Toma de Decisiones Basada en Datos
 Objetivo: Utilizar datos para informar decisiones empresariales, generando insights claros y comunicando resultados de manera efectiva.
 Contenido:
 1. Contando historias con datos (Data Storytelling)
 2. Análisis de datos para la toma de decisiones empresariales.
 3. Cómo generar Hallazgos a partir de los resultados de los análisis.
 4. Comunicación de resultados y decisiones informadas.
14. Tema 14: Proyecto Final - Desarrollo y Asesoría
 Objetivo: Aplicar todos los conocimientos adquiridos en el curso para desarrollar un proyecto completo de Ciencia de Datos, recibiendo retroalimentación para mejoras.
 Contenido:
 1. Asesoría para el desarrollo de proyectos finales.
 2. Revisión y retroalimentación de avances.

Bibliografía básica
Nield, T. (2022). Essential math for data science: Take control of your data with fundamental linear algebra, probability, and statistics. O'Reilly Media. ISBN: 978-1-098-10293-7.
Bruce, P., Bruce, A., & Gedeck, P. (2022). Estadística práctica para ciencia de datos con R y Python. Marcombo. ISBN: 978-84-267-3443-3.
Nussbaumer Knaflic, C. (2017). Storytelling con datos: Visualización de datos para profesionales. Anaya Multimedia. ISBN: 978-84-415-3930-3.
Grus, J. (2019). Data science from scratch: First principles with Python (2nd ed.). O'Reilly Media. ISBN: 978-1-492-04113-9

Bibliografía complementaria
Matthes, E. (2019). Python Crash Course: A Hands-On, Project-Based Introduction to Programming (2nd ed.). No Starch Press.
Spiegelhalter, D. J. (2019). The Art of Statistics: How to Learn from Data. Basic Books.
Tavarez, A. (2021). Ciencia de Datos: Una Guía Práctica. Independently Published.

Materiales del curso
Cuenta activa de Google (Gmail, Google Colab)
Acceso a internet

Evaluación	
Proyecto	60%
Exámenes	10%
Prácticas	30%

Información del profesor	
Nombre completo:	Maestro Carlos Saucedo
Correo electrónico institucional:	carlos.saucedo@ingenieria.unam.edu
Horario de la clase:	Martes y Jueves de 17:00 A 18.30 HRS
Semblanza corta del profesor.	Director Asociado, Impacto Social, Latinoamérica, Kyndryl.