

# “INTRODUCCIÓN A LA CIENCIA DE DATOS”

## DATOS GENERALES

Tipo de crédito	Tipo de asignatura	Idioma de impartición	Modalidad de impartición
Optativa	Curso	Español	Presencial y/o Mixta

## CRÉDITOS

De acuerdo con la propuesta curricular, los datos escolares de la asignatura son:

Semestre	Número de semanas	Horas presenciales de teoría por semana	Horas presenciales de práctica por semana	Horas de trabajo autónomo del estudiante por semana	Total de créditos (RGEP)
Optativa	16	3	0	5	8

## OBJETIVO GENERAL DE APRENDIZAJE

Introducir al estudiante de posgrado a la ciencia de datos partiendo de conceptos básicos de matemáticas y programación, hasta el uso de software especializado como Python, Jupyter Notebooks y R; colocando un énfasis particular en la recolección, procesamiento, análisis, visualización y manejo de bases de datos.

## COMPETENCIAS PROFESIONALES A LAS QUE CONTRIBUYE LA ASIGNATURA

Esta asignatura contribuye de manera directa al logro de las siguientes competencias profesionales del perfil de egreso del programa:

Competencia	Descripción de la competencia
<b>Transversal</b>	Resolver problemas en alguna de las diferentes líneas de generación y aplicación del conocimiento del programa de maestría (Neurociencias, Biología Funcional, Ecología Integrativa y Conservación, y Bioingeniería) mediante el uso de metodologías y herramientas biológicas, analíticas y de ingeniería con énfasis en salud, conservación y medio ambiente, buscando contribuir al desarrollo de biotecnologías.
<b>Profesional específica</b>	Realizar análisis y modelado de información aplicados a sistemas biológicos y médicos, con base a herramientas de ciencia de datos.
<b>Profesional de énfasis</b>	Realizar actividades de investigación y/o desarrollo tecnológico, solucionar problemas de amplio impacto social, con una perspectiva multidisciplinar, en áreas y disciplinas asociadas a la Fisiología, Biología Molecular, Genética, Biología Celular, Microbiología, Ecología y Conservación del Ambiente, Conservación, Neurociencias, y Bioingeniería, o en cualquiera de las áreas de especialización del posgrado.

## PLANEACIÓN DIDÁCTICA GENERAL

A continuación, se describe la planeación general del proceso de aprendizaje:

#	Nombre de la Unidad o Fase	Resultados de aprendizaje específicos	Metodologías y actividades de enseñanza-aprendizaje
1	Introducción 1.1 Concepto de la ciencia de datos y sus relaciones con otros campos de la inteligencia artificial 1.2 Impacto de la ciencia de datos en la sociedad	Proporcionar un panorama general de la ciencia de datos y sus aplicaciones e impacto.	<ul style="list-style-type: none"> <li>Revisar las definiciones existentes en la literatura acerca de la ciencia de datos.</li> <li>Evaluar la relación entre la ciencia de datos con la estadística, la minería de datos, el aprendizaje de máquina, y los grandes datos.</li> <li>Analizar el impacto y consecuencias de la ciencia de datos en la industria y la sociedad.</li> <li>Investigar el plan curricular y el perfil de egreso de un programa de licenciatura enfocado en ciencia de datos en México, y otro en Estados Unidos o Europa, así como de un programa de maestría en México, y otro en Estados Unidos o Europa.</li> </ul>
2	Herramientas para la Ciencia de Datos 2.1 Lenguaje de programación Python 2.2 Librerías de Python 2.3 Ambientes de desarrollo en Python 2.4 Manejo de bases de datos en Python	Introducir al entorno de programación de Python, y su aplicación para la lectura, procesamiento, y visualización de bases de datos.	<ul style="list-style-type: none"> <li>Introducir al alumno al lenguaje Python y su filosofía de programación</li> <li>Revisar las librerías de Python que se utilizan comúnmente en ciencia de datos: Numpy, Pandas, Scipy y Scikit Learn</li> <li>Conocer los ambientes integrados de desarrollo para Python: Spyder, Jupyter Notebook, y Colab.</li> <li>Introducir al alumno a las bases de programación en Python: tipo de variables, condicionales y ciclos.</li> <li>Revisar los comandos principales en Python para la manipulación de bases de datos con la librería Pandas</li> <li>Realizar la exploración de una base de datos de acceso libre con Python</li> </ul>
3	Conceptos Básicos de Estadística 3.1 Recolección y exploración de bases de datos 3.2 Estimación de parámetros 3.3 Estadística inferencial 3.4 Variabilidad de una estimación 3.5 Prueba de hipótesis	Estudiar los conceptos básicos de estadística descriptiva e inferencial asociados con el análisis y estudio de bases de datos.	<ul style="list-style-type: none"> <li>Revisar las técnicas de recolección y comandos en Python para la exploración de una base de datos</li> <li>Introducir al alumno a la estimación de parámetros descriptivos de una base de datos</li> <li>Estudiar el problema de estimación y cuantificar su variabilidad</li> <li>Definir el concepto y los problemas a estudiar en la estadística inferencial</li> <li>Plantear el concepto y las técnicas para una prueba de hipótesis</li> <li>Realizar la exploración estadística de una base de datos de acceso libre con Python</li> </ul>
4	Aprendizaje Supervisado 4.1 Concepto de aprendizaje	Definir el concepto de aprendizaje supervisado, y las técnicas para su	<ul style="list-style-type: none"> <li>Definir el problema de aprendizaje en la ciencia de datos y sus implicaciones</li> <li>Revisar los conceptos de complejidad y sobre-ajuste durante el proceso de aprendizaje</li> </ul>

	4.2 Curvas de aprendizaje 4.3 Entrenamiento, validación y evaluación 4.3 Modelos de aprendizaje	implementación y evaluación.	<ul style="list-style-type: none"> <li>Introducir al alumno a identificar el propósito de las etapas de entrenamiento, validación y evaluación</li> <li>Estudiar tres modelos de aprendizaje elementales: máquinas de soporte vectorial, árboles de decisión, bosques aleatorios, y redes neuronales</li> <li>Estudiar las técnicas de comparación entre modelos</li> <li>Entrenar un modelo de aprendizaje supervisado con una base de datos de acceso abierto para un problema de clasificación</li> </ul>
5	Análisis de Regresión 5.1 Regresión lineal y polinomial 5.2 Regresión multivariada y regularizada 5.3 Regularización con modelos de aprendizaje 5.4 Regresión logística	Desarrollar el concepto de regresión y su aplicación en modelos de predicción.	<ul style="list-style-type: none"> <li>Repasar el concepto y las técnicas de regresión lineal</li> <li>Estudiar el planteamiento de la regresión polinomial y su solución</li> <li>Definir las bases y algoritmos para la regresión multivariada</li> <li>Estudiar los modelos de regresión regularizada</li> <li>Revisar las técnicas de regresión basadas en modelos de aprendizaje</li> <li>Estudiar las bases de la regresión logística para un escenario de clasificación</li> <li>Emplear las técnicas de regresión para un problema de predicción con una base de datos de acceso abierto</li> </ul>
6	Aprendizaje No-supervisado 6.1 Definición del aprendizaje no-supervisado 6.2 Análisis de cúmulos 6.3 Reducción de dimensionalidad	Definir el concepto de aprendizaje no supervisado, y el uso del análisis de cúmulos con este fin.	<ul style="list-style-type: none"> <li>Introducir a los alumnos al concepto de aprendizaje no-supervisado</li> <li>Estudiar las técnicas básicas en la agrupación por cúmulos</li> <li>Aplicar las técnicas de agrupación a un caso de estudio</li> <li>Estudiar las bases teóricas y algoritmos para la reducción de dimensionalidad</li> <li>Utilizar las técnicas de reducción de dimensionalidad en una base de datos de acceso abierto para entrenar un modelo de aprendizaje</li> </ul>

## EVALUACIÓN

A continuación, se muestra las condiciones de las evaluaciones parciales.

# Parcial	Momento de evaluación	Método de evaluación y valor para la evaluación parcial	Ponderación para evaluación final
1	Al final de la unidad 1	<ul style="list-style-type: none"> <li>Asignación de un proyecto de investigación acerca de programas de estudio en ciencia de datos</li> </ul>	5%
2	Al final de la unidad 2	<ul style="list-style-type: none"> <li>Asignación de un proyecto práctico en Python para la exploración de una base de datos</li> </ul>	10%
3	Al final de la unidad 3	<ul style="list-style-type: none"> <li>Asignación de un proyecto práctico en Python para el análisis estadístico de una base de datos</li> </ul>	10%

4	Al final de la unidad 4	<ul style="list-style-type: none"> <li>Asignación de un proyecto práctico en Python para entrenar modelos de aprendizaje supervisado con una base de datos de acceso abierto para un problema de clasificación</li> </ul>	15%
5	Al final de la unidad 5	<ul style="list-style-type: none"> <li>Asignación de un proyecto práctico en Python para emplear las técnicas de regresión en un problema de predicción con una base de datos de acceso abierto</li> </ul>	15%
6	Al final de la unidad 6	<ul style="list-style-type: none"> <li>Asignación de un proyecto práctico en Python para utilizar las técnicas de reducción de dimensionalidad en una base de datos de acceso abierto para entrenar un modelo de aprendizaje</li> </ul>	15%
7	Al concluir todas las unidades del curso	<ul style="list-style-type: none"> <li>Asignación de un proyecto final asociado a la aplicación de la ciencia de datos en la línea de investigación del alumno y que involucre el uso de Python</li> </ul>	30%

## RECURSOS BIBLIOGRÁFICOS Y DIGITALES

### TEXTOS BÁSICOS

- “Introduction to Data Science: A Python Approach to Concepts, Techniques and Applications”, Laura Igual y Santi Seguí, Springer-Verlag Ed., 2017.
- “Data Science from Scratch”, Joel Grus, 2<sup>nd</sup> Ed, Ed. O’Reilly, 2019.
- “Introducing Data Science”, D. Cielen, A.D.B. Meysman y Mohamed Ali, Manning Publications Co., 2016.
- “R for Data Science”, Hadley Wickham y Garrett Grolemund, Ed. O’Reilly, 2017.
- “Python Crash Course”, Eric Matthes, No Starch Press, 2<sup>nd</sup> Ed, 2019.

### RECURSOS DIGITALES

- ORACLE <https://www.oracle.com/mx/what-is-data-science/>
- AWS <https://aws.amazon.com/es/what-is/data-science/>
- IBM <https://www.ibm.com/analytics/mx/es/technology/data-science/>

## REQUISITOS PARA CURSAR LA ASIGNATURA

Para poder cursar esta asignatura, es necesario acreditar inicialmente:

- Diseño de Investigación y Bioestadística

## INTEROPERABILIDAD

Esta asignatura es compartida con los siguientes programas de posgrado:

- Maestría en Ingeniería Electrónica
- Doctorado en Ciencias de la Ingeniería
- Maestría y Doctorado en Ciencias Aplicadas

## OTRAS FORMAS DE ACREDITACIÓN

- Esta asignatura puede ser acreditada a través de la presentación de un documento probatorio que certifique que el estudiante ya cuenta con los aprendizajes necesarios: **Si**
- Esta asignatura puede ser acreditada a través de un examen que certifique que el estudiante ya cuenta con los aprendizajes necesarios: **Si**.

## MÁXIMO Y MÍNIMO DE ESTUDIANTES POR GRUPO

- Máximo de estudiantes por grupo para garantizar viabilidad académica, pedagógica y financiera: 20
- Mínimo de estudiantes por grupo para garantizar viabilidad académica, pedagógica y financiera: 3

## ELABORADORES Y REVISORES

- **Elaboró:** Dr. Daniel Ulises Campos Delgado
- **Revisó:**