



INFORMACIÓN ESPACIO ACADÉMICO						
Nombre de asignatura	CIENCIA DE DATOS					
Código	11501003					
Definición de asignatura	Obligatorio	X	Básico		Complementario	
	Electivo		Intrínseco	X	Extrínseco	
Número de créditos	3	Horas	144	Semanas	16	
Distribución créditos	HTD	64	HTC	32	HTA	48
	Hora de trabajo cooperativo					
Metodología	Virtual					

PROGRAMACIÓN DEL CONTENIDO
<b>CONOCIMIENTOS PREVIOS</b>
<ul style="list-style-type: none"> <li>Tecnologías de la Información (TI)</li> <li>Estadística Básica</li> </ul>
<b>DESCRIPCIÓN DEL CURSO</b>
<p>Este curso de Ciencia de Datos busca desarrollar competencias sólidas en técnicas analíticas y de evaluación de datos. Se busca generar un entendimiento de los fundamentos y aplicaciones prácticas en recolección, procesamiento, visualización, y análisis de datos, incluyendo métodos estadísticos y técnicas de aprendizaje automático y minería de datos. El curso introduce conceptos cruciales como la preparación y limpieza de datos, la construcción y evaluación de modelos de aprendizaje automático supervisados y no supervisados, y técnicas de minería de datos para descubrir patrones ocultos y conocimientos valiosos. Se abordará el manejo y análisis de Big Data, y se explorarán herramientas y plataformas pertinentes. Finalmente, los estudiantes aprenderán a presentar sus hallazgos mediante técnicas efectivas de visualización de datos. En general, el objetivo es inculcar una mentalidad de resolución de problemas y toma de decisiones basada en datos, equipando a los estudiantes para manejar los desafíos de la ciencia de datos en situaciones reales.</p>
<b>ÁREAS DE CONOCIMIENTO</b>
<ul style="list-style-type: none"> <li>Estadística descriptiva</li> <li>Minería de Datos</li> <li>Aprendizaje automático</li> <li>Big Data</li> </ul>
<b>COMPETENCIAS EN FORMACIÓN</b>
<p>Al final de la asignatura el estudiante podrá contar con las siguientes competencias:</p> <ul style="list-style-type: none"> <li>Los estudiantes deberán desarrollar habilidades para recolectar, limpiar, transformar y preparar datos para análisis. Deben entender la importancia de la calidad de los datos y cómo afecta los resultados del análisis.</li> <li>Los estudiantes deberán entender y aplicar técnicas estadísticas para analizar y explorar datos, incluyendo la inferencia estadística, análisis de regresión y medidas estadísticas.</li> <li>Los estudiantes deberán desarrollar un entendimiento sólido de los algoritmos de aprendizaje automático, tanto supervisados como no supervisados, y cómo estos pueden</li> </ul>





ser aplicados para realizar predicciones y descubrir patrones ocultos en los datos.

- Los estudiantes deberán ser capaces de aplicar técnicas de minería de datos para extraer información valiosa y conocimientos a partir de grandes conjuntos de datos.
- Los estudiantes deberán ser capaces de manejar y analizar grandes volúmenes de datos, utilizando herramientas y plataformas de Big Data como Spark y Pandas.
- Los estudiantes deberán desarrollar habilidades para presentar datos y hallazgos de análisis de manera efectiva utilizando técnicas de visualización de datos.
- Los estudiantes deberán demostrar una mentalidad de resolución de problemas, siendo capaces de aplicar los principios y técnicas de la ciencia de datos para resolver problemas complejos y tomar decisiones basadas en datos en situaciones del mundo real.

## ESTRUCTURA DEL MÓDULO

### UNIDAD 1. Ciencia de Datos y Manejo de Datos

OBJETIVO DE APRENDIZAJE	TÍTULO DE TEMA
	1.1 Desarrollo, Historia y perspectivas.
	1.2 El proceso de ciencia de datos: Recolección, preprocesamiento y limpieza de datos.
	1.3 Fundamentos de programación en Python y R para ciencia de datos.
	1.4 Manipulación de datos con pandas y data.table.

### UNIDAD 2. Estadísticas y Análisis Exploratorio de Datos

OBJETIVO DE APRENDIZAJE	TÍTULO DE TEMA
	2.1 Fundamentos de estadística: media, mediana, desviación estándar (tendencia central y dispersión)
	2.2 Pruebas estadísticas: t-test, chi-cuadrado, ANOVA.
	2.3 Regresión lineal y logística.
	2.4 Visualización de datos para análisis exploratorio: matplotlib, ggplot2.

### UNIDAD 3. Aprendizaje Automático

OBJETIVO DE APRENDIZAJE	TÍTULO DE TEMA
	3.1 aprendizaje automático: supervisado, no supervisado, semi-supervisado.
	3.2 Modelos de regresión y clasificación: regresión lineal, árboles de decisión, SVM
	3.3 Modelos de agrupación: k-means, DBSCAN.
	3.4 Reducción de la dimensionalidad y técnicas de incrustación: PCA, t-SNE.

### UNIDAD 4. Minería de Datos y Descubrimiento de Conocimientos

OBJETIVO DE APRENDIZAJE	TÍTULO DE TEMA
	4.1 Introducción a la minería de datos: procesos y técnicas.
	4.2 Modelos predictivos y técnicas de clasificación.
	4.3 Asociación y técnicas de agrupación.
	4.4 Minería de texto y análisis de sentimiento.

### UNIDAD 5. Big Data y Análisis de Datos a Gran Escala

OBJETIVO DE APRENDIZAJE	TÍTULO DE TEMA





	5.1 Introducción a Big Data: Conceptos y desafíos.
	5.2 Infraestructura de Big Data: Hadoop y Spark.
	5.3 Bases de datos NoSQL para manejo de Big Data: MongoDB, Cassandra.
	5.4 Aplicaciones y casos de estudio en Big Data.

Tenga en cuenta las siguientes **estrategias de aprendizaje** para el planteamiento de las actividades de evaluación:

- **Estudio autónomo:** lectura y revisión de las unidades, de los recursos tales como videos, lecturas, hipervínculos, investigación, exploración en redes académicas
- **Tutoría:** revisión de clases magistrales virtuales, asistencia a tutoría virtuales presenciales, comunicación con el tutor y aclaración de dudas.
- **Autoevaluaciones:** cuestionarios de evaluación en línea
- **Prácticas:** actividades durante el desarrollo del curso de diferente índole, orientadas a proyectos, problemas, investigación, estudio de caso, entre otras
- **Trabajo final:** elaboración de una actividad que integra lo desarrollado durante el curso, la cual se debe entregar la última semana del curso.
- **Notas:** las actividades se pueden desarrollar tanto individual como grupal, según criterio del docente.

EVALUACIÓN		
TIPO	EVALUACIÓN/ACTIVIDAD	PORCENTAJE
Continua	Actividad Inicial (Foro), Cuadro comparativo de Metodologías de Gestión de Riesgos de Ciberseguridad, Fase 1 Proyecto Final.	35%
Formativa	Parcial virtual Unidad 1, mapa conceptual controles sistemas de información, fase 2 proyecto final	35%
Formativa	Fase 3 Proyecto Final. Presentación completa del Proyecto Final incluyendo, nivel de madurez, controles a implementar y planeación de la Implementación.	30%
<b>Total del curso</b>		<b>100 %</b>

BIBLIOGRAFÍA	
1	Provost, F., & Fawcett, T. (2013). Data Science for Business: What you need to know about data mining and data-analytic thinking. O'Reilly Media, Inc.
2	James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). An Introduction to Statistical Learning. Springer.





3	Hastie, T., Tibshirani, R., & Friedman, J. (2009). The Elements of Statistical Learning: Data Mining, Inference, and Prediction. Springer.
4	McKinney, W. (2012). Python for Data Analysis: Data Wrangling with Pandas, NumPy, and IPython. O'Reilly Media, Inc.
5	Bishop, C. M. (2006). Pattern Recognition and Machine Learning. Springer
6	Wickham, H., & Golemund, G. (2016). R for Data Science: Import, Tidy, Transform, Visualize, and Model Data. O'Reilly Media, Inc.
7	Kelleher, J. D., Mac Namee, B., & D'Arcy, A. (2020). Fundamentals of Machine Learning for Predictive Data Analytics: Algorithms, Worked Examples, and Case Studies. MIT Press.
8	Chollet, F., & Allaire, J. J. (2018). Deep Learning with R. Manning Publications.
9	Stal Grus, J. (2018). Data Science from Scratch: First Principles with Python. O'Reilly Media, Inc.

PROGRAMA SINTÉTICO		ORGANIZACIÓN / TIEMPOS															
		SEMANAS ACADÉMICAS															
		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
1.	Ciencia de Datos y Manejo de Datos	X	X														
2.	Estadísticas y Análisis Exploratorio de Datos			X	X	X	X										
3.	Aprendizaje Automático							X	X	X	X						
4.	Minería de Datos y Descubrimiento de Conocimientos											X	X	X			
5.	Big Data y Análisis de Datos a Gran Escala														X	X	X

ELABORÓ: Azeneth Rozo Dicelis



Esta obra está bajo una licencia: **CC BY-NC-ND**

Atribución – No comercial – Sin derivar

Consultar información relacionada en: [Atribución – No comercial – Sin derivar](#)