



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

PROGRAMACIÓN DEL CONTENIDO
CONOCIMIENTOS PREVIOS
<ul style="list-style-type: none"> Tecnologías de la Información (TI) Estadística Básica
DESCRIPCIÓN DEL CURSO
<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>
ÁREAS DE CONOCIMIENTO
<ul style="list-style-type: none"> Estadística descriptiva Minería de Datos Aprendizaje automático Big Data
COMPETENCIAS EN FORMACIÓN
<p>Al final de la asignatura el estudiante podrá contar con las siguientes competencias:</p> <ul style="list-style-type: none"> Conocer y utilizar las técnicas y herramientas para recolectar, limpiar, transformar y preparar datos para análisis, incluyendo los aspectos de calidad de los datos. Perfeccionar el entendimiento en técnicas estadísticas utilizadas para los esquemas analíticos, al igual que en los modelos y algoritmos de aprendizaje automático. Desarrollar habilidades para construir modelos de aprendizaje de máquina que se integren dentro del esquema de datos, con capacidades para ajustar parámetros de entrenamiento.





- Fortalecer las capacidades analíticas y computacionales para manejar grandes volúmenes de datos, utilizando herramientas ampliamente utilizadas por la industria y la comunidad.

ESTRUCTURA DEL MÓDULO

UNIDAD 1.

OBJETIVO DE APRENDIZAJE	Introducción a la ciencia e Ingeniería de Datos
Desarrollar las habilidades básicas de análisis, manipulación y programación para ciencia de datos.	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.

OBJETIVO DE APRENDIZAJE	Estadística y Análisis Exploratorio de Datos
Fortalecer las bases de estadística descriptiva y desarrollar capacidades para ejecutar operaciones estadísticas que permitan realizar predicciones de valores futuros.	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.

OBJETIVO DE APRENDIZAJE	Aprendizaje Automático y Minería de Datos
Crear las capacidades básicas para entrenar modelos de aprendizaje automático y minería de datos utilizando las librerías y lenguajes más populares en ciencia de datos.	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.
	3.5. introducción a la minería de datos: procesos y técnicas.
	3.6 Modelos predictivos y técnicas de clasificación.
	3.7 Asociación y técnicas de agrupación.
	3.8 Minería de texto y análisis de sentimiento.

UNIDAD 4.

OBJETIVO DE APRENDIZAJE	Big Data y Análisis de Datos a Gran Escala
Utilizar las herramientas analíticas	4.1 Introducción a Big Data: Conceptos y desafíos.





para resolver problemas y conjunto de datos en un ambiente de gran escala.	4.2 Infraestructura de Big Data: Hadoop y Spark.
	4.3 Bases de datos NoSQL para manejo de Big Data: MongoDB, Cassandra.
	4.4 Aplicaciones y casos de estudio en Big Data.

EVALUACIÓN		
TIPO	EVALUACIÓN/ACTIVIDAD	PORCENTAJE
Formativa	Foros: Se contemplan tres (3) foros en los que se anima a la discusión y participación de temas técnicos en un escenario real y se reflexiona sobre las implicaciones de los mismos. En estos foros se tratan temas generales sobre los datos, los modelos de aprendizaje de máquina y NLP y los desafíos al tratar con grandes volúmenes de datos	10%
Formativa	Trabajo Colaborativo. Se anima al trabajo en grupo para responder a desafíos relacionados con tópicos técnicos orientados a resolver problemas de la actualidad.	5%
Continuo	Laboratorios: Se desarrolla actividades de programación en Python y se encadena en las diferentes etapas de los datos, Los resultados se verán reflejados en un notebook donde se muestran y comunican las conclusiones.	60%
Formativa	Ensayo. Se establece un (1) ensayo individual sobre las implicaciones de la fundamentación matemática en las técnicas modernas de análisis de datos.	0%
Formativa	Proyecto final. Este proyecto será construido de manera gradual basado en los laboratorios planteados a lo largo de las unidades y en cada una de ellas se agregará complejidad de tal manera que al final será un proyecto terminado y con presentación de resultados.	25%
Total del curso		100 %

BIBLIOGRAFÍA	
1	Reis, J., Housley, M. (2022). Fundamentals of Data Engineering. Estados Unidos: O'Reilly Media.





2	James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). An Introduction to Statistical Learning. Springer. disponible en https://www.statlearning.com/
3	Diez, D. M., Barr, C., & Çetinkaya-Rundel, M. (2022). OpenIntro Statistics: A new standard for high school and college (4th ed.). OpenIntro, Inc. Disponible en https://www.openintro.org/book/os/
4	VanderPlas, J. (2022). Python Data Science Handbook: Essential Tools for Working with Data. O'Reilly Media.
5	Müller, A. C., & Guido, S. (2021). Introduction to Machine Learning with Python: A Guide for Data Scientists. O'Reilly Media.
6	Ivan Marin Ankit Shukla , Sarang VK(2019). Big Data Analysis with Python: Combine Spark and Python to unlock the powers of parallel computing and machine learning. Apress
7	Damji, J. S., Wenig, B., Lee, D., Das, T. (2020). Learning Spark. Estados Unidos: O'Reilly Media
8	Géron, A. (2020). Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems (2nd ed.). O'Reilly Media.

ORGANIZACIÓN / TIEMPOS																
PROGRAMA SINTÉTICO	SEMANAS ACADÉMICAS															
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
1. Introducción a la ciencia e Ingeniería de Datos	X	X	X													
2. Estadística y Análisis Exploratorio de Datos				X	X	X	X									
3. Aprendizaje Automático y Minería de Datos								X	X	X	X	X	X			
4. Big Data y Análisis de Datos a Gran Escala														X	X	X

ELABORÓ: Azeneth Rozo Dicelis

