

IDENTIFICACIÓN DE LA UNIDAD DE APRENDIZAJE

Unidad académica: Instituto de Investigación en Ciencias Básicas y Aplicadas.							
Plan de estudios: Licenciatura en Física y Matemáticas							
Unidad de aprendizaje: Introducción a la ciencia de datos				Ciclo de formación: Especializado Eje general de formación: Teórico-técnica Área de conocimiento: Matemáticas avanzada Semestre: 8°			
Elaborada por: Dr. Daniel Rivera López				Fecha de elaboración: Marzo, 2021			
Clave:	Horas teóricas	Horas prácticas:	Horas totales	Créditos	Tipo de unidad de aprendizaje	Carácter de la unidad de aprendizaje:	Modalidad:
OPP38CE050010	5	0	5	10	Optativa	Teórica	Escolarizada
Programa Educativo en el que se imparte: Licenciatura en Física y Matemáticas del Instituto de Investigación en Ciencias Básicas y Aplicadas.							

ESTRUCTURA DE LA UNIDAD DE APRENDIZAJE

Presentación: La ciencia de datos es un campo interdisciplinario que involucra métodos científicos, procesos y sistemas para extraer conocimiento o un mejor entendimiento de datos en sus diferentes formas, ya sea estructurados o no estructurados. Esta unidad de aprendizaje tendrá enfoque teórico-práctico haciendo énfasis en la práctica. Los contenidos teóricos tienen como objetivo exponer las ideas matemáticas y computacionales sin entrar en detalles de rigor técnico. Los contenidos prácticos se centrarán en casos de estudio, simulaciones por computadora, prácticas de codificación en el lenguaje de programación Python, así como proyectos.
Propósito: Busque, analice y utilice datos relevantes de un tema de investigación, mediante la aplicación de técnicas de análisis estadístico, exploración y análisis, la formulación de preguntas relevantes, narrativas y visualizaciones, con el fin de guiar la toma de decisiones; en particular clasificar, predecir fenómenos y comunicar resultados de manera eficaz con compromiso ético y con la calidad.
Competencias que contribuyen al perfil de egreso.
Competencias genéricas:
CG4. Capacidad de abstracción, análisis y síntesis. CG11. Habilidades para buscar, procesar y analizar información. CG12. Habilidad para el trabajo en forma colaborativa. CG21. Capacidad de expresión y comunicación. CG32. Compromiso con la calidad. CG33. Compromiso ético.
Competencias específicas:
CE 2. Formula problemas en lenguaje matemático y contribuye a la construcción de modelos matemáticos, mediante la aplicación de teorías, fórmulas y principios matemáticos, con el fin de facilitar su análisis y solución en los sectores públicos, privados o sociales con rigor metódico, precisión y certeza.

CE 3. Utiliza y diseña programas o sistemas de computación mediante el uso de equipo especializado, para el procesamiento de información, cálculo numérico y simulación de procesos que permitan dar soluciones innovadoras a problemas planteados con objetividad y responsabilidad.

CE 5. Posee conocimientos, habilidades, valores y actitudes requeridos en investigación inter y multidisciplinaria de las ciencias básicas y aplicadas, mediante el análisis, intercambio y producción de información entre grupos académicos de diferentes campos disciplinares que involucren a la física y la matemática, para contribuir científicamente en equipos de investigación con un sentido de trabajo colaborativo y profesional.

CE 7. Comunica asertivamente conceptos, objetivos, métodos y resultados del lenguaje científico, mediante la comunicación oral y escrita, para presentar propuestas y proyectos de manera eficaz, funcional y aplicable.

CONTENIDOS

Bloques:	Temas:
I. Qué es la ciencia de datos.	1.1 Introducción histórica y comparación con áreas relacionadas. 1.2 Formulación de preguntas interesantes sobre datos. 1.3 Taxonomía de los datos 1.4 Introducción a los problemas de clasificación y regresión.
II. Herramientas para ciencia de datos.	2.1 Software y formatos de archivo para ciencia de datos. 2.2 Recolección de datos. 2.3 Limpieza de datos 2.4 Puntuaciones y ránkines. 2.5 Reducción de dimensionalidad
III. Análisis exploratorio de datos.	3.1 Estadística descriptiva y análisis de correlación 3.2 Visualización de datos, sus valores estéticos y buenas prácticas. 3.3 Narración de historias con datos. 3.4 Ejemplos de visualizaciones interesantes.
IV. Análisis estadístico.	4.1 Distribuciones de probabilidades comunes. 4.2 Muestreo de distribuciones 4.3 Significación estadística 4.4 Pruebas de permutación y p-valores. 4.5 Introducción al razonamiento bayesiano.
V. Introducción al aprendizaje automático.	5.1 Evaluación de modelos 5.2 Regresión lineal y logística 5.3 El método de los k-vecinos más cercanos 5.4 Árboles de decisión y bosques aleatorios 5.5 Máquinas de soporte vectorial 5.6 Agrupamiento por k-medias
VI. Tópicos selectos.	6.1 Procesamiento de grandes volúmenes de datos. 6.2 Procesamiento de lenguaje natural. 6.3 Sistemas de recomendación. 6.4 Análisis de series de tiempo. 6.5 Redes neuronales artificiales y análisis profundo.

ESTRATEGIAS DE ENSEÑANZA - APRENDIZAJE

Estrategias de aprendizaje sugeridas (Marque X)			
Aprendizaje basado en problemas	(x)	Nemotecnia	()
Estudios de caso	(x)	Análisis de textos	()
Trabajo colaborativo	()	Seminarios	()
Plenaria	()	Debate	()
Ensayo	()	Taller	(x)
Mapas conceptuales	()	Ponencia científica	()
Diseño de proyectos	()	Elaboración de síntesis	()
Mapa mental	()	Monografía	()
Práctica reflexiva	()	Reporte de lectura	()
Trípticos	()	Exposición oral	()
Otros			
Estrategias de enseñanza sugeridas (Marque X)			
Presentación oral (conferencia o exposición) por parte del profesorado	(x)	Experimentación (prácticas)	()
Debate o Panel	()	Trabajos de investigación documental	(x)
Lectura comentada	(x)	Anteproyectos de investigación	()
Seminario de investigación	()	Discusión guiada	(x)
Estudio de Casos	(x)	Organizadores gráficos (Diagramas, etc.)	()
Foro	()	Actividad focal	()
Demostraciones	()	Analogías	()
Ejercicios prácticos (series de problemas)	(x)	Método de proyectos	()
Interacción la realidad (a través de videos, fotografías, dibujos y software especialmente diseñado).	(x)	Actividades generadoras de información previa	()
Organizadores previos	()	Exploración de la web	()
Archivo	()	Portafolio de evidencias	()
Ambiente virtual (foros, chat, correos, ligas a otros sitios web, otros)	()	Enunciado de objetivo o intenciones	()
Otra, especifique (lluvia de ideas, mesa redonda, textos programados, cine, teatro, juego de roles, experiencia estructurada, diario reflexivo, entre otras):			

CRITERIOS DE EVALUACIÓN

Criterios sugeridos	Porcentaje
• Exámenes parciales	30%
• Examen final	30%
• Participación en clase	10%
• Tareas	30%
Nota: Algunos de los instrumentos de evaluación que se pueden considerar son: Rúbricas, escalas de cotejo, escala estimativa, entre otros.	
Total	100 %

PERFIL DEL PROFESORADO

Preferentemente con nivel Doctorado en Física, Matemáticas o área afín a la disciplina de la unidad de aprendizaje, que asegure un dominio integral de los saberes en su campo, es deseable que cuente con experiencia docente y en la generación y aplicación del conocimiento como ejercicio de su profesión.

REFERENCIAS

Básicas:

- Laura Igual, Santi Seguí. (2017). Introduction to Data Science: A Python Approach to Concepts, Techniques and Applications. Springer.
- Sinan Ozdemir, Sunil Kakade and Marco Tibaldeschi. (2018). Principles of Data Science, Second Edition. Packt, 2018.
- Steven, S. Switzerland (2017). Data science design manual. Springer.

Complementarias:

- Aggarwal, C. C. Cham (2016). Data mining: The textbook. Springer.
- Donoho, D. 50 Years of Data Science. Journal of Computational and Graphical Statistics, 26:4, 745-766, 2017.
- Provos Foster, Fawcett Tom (2013). Data Science for Business What you need to know about data mining and data-analytic thinking. Köln: O`Reilly.