
5. CIENCIA DE DATOS

1. Datos Generales de la asignatura

Nombre de la asignatura:	Ciencia de datos
Clave de la asignatura:	TNB-2305
SATCA⁵:	2 – 3 – 5
Carrera:	Ingeniería en Sistemas Computacionales

2. Presentación

Caracterización de la asignatura
<p>Esta asignatura aporta al perfil del egresado la capacidad para procesar, interpretar y graficar datos.</p> <p>La importancia de esta materia radica en que el estudiante podrá aplicar sus conocimientos y habilidades para análisis de grandes cantidades de datos.</p> <p>La materia consiste en cuatro partes: la primera explica los conceptos básicos sobre Ciencia de Datos. La segunda se enfoca en Probabilidad y Estadística aplicada a Ciencia de Datos. La tercera es sobre Machine Learning y Deep Learning las cuales aportan nuevos algoritmos y formas de procesar la información. La última unidad se encarga del Procesamiento del Lenguaje Natural tan útil en nuestros días para un sinfín de aplicaciones.</p> <p>Esta materia es de las últimas del módulo de especialidad y en ella se involucran la mayoría de los conocimientos adquiridos durante toda la carrera.</p>
Intención didáctica
<p>El profesor deberá contar con experiencia en el área de Ciencia de Datos y haber participado en proyectos relacionados con esta área.</p> <p>Deberá desarrollar la capacidad para coordinar el trabajo en equipo, así como proponer actividades para el aprendizaje significativo que consideren los distintos estilos de aprendizaje de los estudiantes, el entorno de la institución, la formación del profesor y el</p>

ámbito profesional en el que se desenvolverán los futuros profesionistas; todo esto con el compromiso de lograr las competencias requeridas al término de la materia.

El temario consiste de 4 unidades. Cada unidad es un tema diferente al anterior y deberán ser vistos en secuencia para poder llevar a cabo un proyecto completo de Ciencia de Datos.

El estudiante deberá conocer la forma encontrar los datos, extraerlos, organizarlos, almacenarlos y aplicar algoritmos para resolver un problema ó hacer predicciones acerca de los datos analizados.

Durante el semestre el alumno conocerá la teoría y se prepara de manera práctica para la realización de un proyecto completo de Ciencia de Datos.

El docente debe conocer los temas de manera profunda y deberá proponer un proyecto semestral que los alumnos desarrollen durante todo el curso.

3. Participantes en el diseño y seguimiento curricular del programa

Lugar y fecha de elaboración o revisión	Participantes	Observaciones
Instituto Tecnológico de Morelia Fecha: Marzo del 2022	Miembros de la LGAC de “Aplicaciones y Servicios sobre Internet” de la Academia de Ingeniería en Sistemas Computacionales. M.C. Juan Jesús Ruiz Lagunas Dr. Heberto Ferreira Medina M.C. Rogelio Ferreira Escutia M.I. Adrián Núñez Vieyra I.S.C. Kenia Aline Ayala Robles Dr. Juan Carlos Olivares Rojas Ing. Alejandro Amaro Flores	Análisis y diseño curricular de los módulos de especialidad las carreras del departamento de Sistemas y Computación. Plan de estudios de la especialidad 2023

	M.C. Jesús Eduardo Alcaraz Chávez Dr. J. Guadalupe Ramos Díaz M.C. Fernando Villaseñor Béjar	
--	---	--

4. Competencia(s) a desarrollar

Competencia(s) específica(s) de la asignatura
<ul style="list-style-type: none"> • Conocer la ubicación de los repositorios de datos y desarrollar aplicaciones de software para extraer su contenido. • Validar correctamente el contenido de un conjunto de datos. • Conocer y aplicar técnicas de Probabilidad y Estadística para el análisis de datos. • Conocer y aplicar técnicas de Machine Learning y Deep Learning para el análisis de datos. • Conocer y aplicar técnicas de la Programación del Lenguaje Natural aplicado a textos.

5. Competencias previas

<ul style="list-style-type: none"> • Diseño de algoritmos y aplicaciones de software, utilizando lenguajes de programación de alto o bajo nivel, para su aplicación en la solución de problemas • Conocer redes de comunicación. • Bases de datos distribuidas. • Probabilidad y Estadística. • Geometría Analítica. • Álgebra Lineal. • Cálculo Diferencial e Integral. • Programación de aplicaciones de escritorio. • Programación Web. • Programación de Apps para dispositivos móviles.
--

6. Temario

No.	Temas	Subtemas
1	Introducción a Ciencia de Datos	1.1 Definición y conceptos 1.2 Historia 1.3 Habilidades de un científico de datos 1.4 Técnicas y herramientas para ciencia de datos 1.5 Origen de los datos 1.6 Formatos y extracción de datos
2	Probabilidad y Estadística	2.1 Población y muestra 2.2 Histogramas, mínimo, máximo y rango 2.3 Media, mediana y moda 2.4 Medidas de dispersión (varianza y desviación estándar) 2.5 Correlación y regresión lineal 2.6 Probabilidad condicional 2.7 Distribuciones de probabilidad 2.8 Modelos gráficos probabilísticos
3	Machine Learning y Deep learning	3.1 Definición y aplicaciones 3.2 Aprendizaje supervisado y no supervisado 3.3 Algoritmos de clasificación 3.4 Support Vector Machines 3.5 Árboles de decisión y Random Forests 3.6 Algoritmos de agrupación 3.7 Redes neuronales 3.8 Redes neuronales profundas (Deep learning) 3.9 Redes neuronales convolucionales 3.10 Generative Adversarial Networks(GANs) 3.11 Aprendizaje por refuerzo
4	Procesamiento del Lenguaje Natural	4.1 Conceptos sobre lenguaje 4.2 Corpus y fuentes léxicas 4.3 Segmentación de textos 4.4 Frecuencias de palabras 4.5 Detección de similitudes 4.6 Análisis de sentimientos 4.7 Análisis de textos completos 4.8 Construcción de bots para lenguaje natural 4.9 Interacción humano-máquina con bots

7. Actividades de aprendizaje de los temas

1. Introducción a Ciencia de Datos	
Competencias	Actividades de aprendizaje
Específica(s): <ul style="list-style-type: none"> Conocer los conceptos básicos 	<ul style="list-style-type: none"> Conocer la definición de Ciencia de Datos. Conocer el desarrollo cronológico

<p>sobre Ciencia de Datos</p> <p>Genéricas:</p> <ul style="list-style-type: none"> • Habilidad en el uso de las Tecnologías de la Información. • Capacidad para identificar, plantear y resolver problemas. • Capacidad para trabajar en equipo interdisciplinario. • Capacidad crítica y autocrítica. • Habilidades interpersonales. • Capacidad de aplicar los conocimientos en la práctica. 	<p>de las diferentes desarrollos de Ciencia de Datos.</p> <ul style="list-style-type: none"> • Conocer los habilidades necesarias que debe tener un científico de Datos. • Conocer las técnicas y herramientas para extracción de datos. • Conocer los diferentes tipos de datos y la forma de extraerlos.
2. Probabilidad y Estadística	
Competencias	Actividades de aprendizaje
<p>Específica(s):</p> <ul style="list-style-type: none"> • Conocer los conceptos de Probabilidad y Estadística aplicados a Ciencia de Datos. <p>Genéricas:</p> <ul style="list-style-type: none"> • Habilidad en el uso de las Tecnologías de la Información. • Capacidad para identificar, plantear y resolver problemas. • Capacidad para trabajar en equipo interdisciplinario. • Capacidad crítica y autocrítica. • Habilidades interpersonales. • Capacidad de aplicar los conocimientos en la práctica. 	<ul style="list-style-type: none"> • Conocer el concepto de población y extraer una muestra representativa. • Conocer las medidas de conteo y dispersión utilizadas en estadística. • Conocer el concepto de correlación y regresión lineal. • Conocer los conceptos básicos de probabilidad. • Conocer las diferentes distribuciones de probabilidad. • Conocer los modelos gráficos probabilísticos.
3. Machine Learning y Deep Learning	
Competencias	Actividades de aprendizaje
<p>Específica(s):</p> <ul style="list-style-type: none"> • Conocer los conceptos de Machine Learning y Deep Learning aplicados a Ciencia de Datos. 	<ul style="list-style-type: none"> • Conocer los conceptos básicos sobre Machine Learning y Deep Learning. • Conocer y aplicar el Aprendizaje Supervisado, el No Supervisado y el Aprendizaje por Refuerzo.

<p>Genéricas:</p> <ul style="list-style-type: none"> • Habilidad en el uso de las Tecnologías de la Información. • Capacidad para identificar, plantear y resolver problemas. • Capacidad para trabajar en equipo interdisciplinario. • Capacidad crítica y autocrítica. • Habilidades interpersonales. • Capacidad de aplicar los conocimientos en la práctica. 	<ul style="list-style-type: none"> • Conocer los diferentes algoritmos de clasificación por agrupamiento. • Conocer las Redes Neuronales. • Conocer las Redes Neuronales Profundas y todos sus diferentes desarrollos derivados de éstas, como las Convolucionales y las GAN's.
4. Procesamiento del lenguaje natural	
Competencias	Actividades de aprendizaje
<p>Específica(s):</p> <ul style="list-style-type: none"> • Conocer el Procesamiento del Lenguaje Natural aplicado a textos. <p>Genéricas:</p> <ul style="list-style-type: none"> • Habilidad en el uso de las Tecnologías de la Información. • Capacidad para identificar, plantear y resolver problemas. • Capacidad para trabajar en equipo interdisciplinario. • Capacidad crítica y autocrítica. • Habilidades interpersonales. • Capacidad de aplicar los conocimientos en la práctica. 	<ul style="list-style-type: none"> • Conocer los conceptos sobre el Lenguaje Natural. • Conocer los Corpus y las fuentes léxicas. • Conocer la forma de extraer, separar y segmentar textos. • Conocer las diferentes técnicas para el análisis de textos. • Conocer y construir chatbots para llevar a cabo una interacción mas compleja entre máquinas y humanos.

8. Prácticas

<ul style="list-style-type: none"> • Leer y procesar archivos de texto planos y en formato CSV. • Conocer la forma extraer datos de un repositorio en Internet y graficar los datos. • Leer y procesar archivos en formato de Excel (XLS). • Leer y procesar archivos en formato JSON. • Leer un archivo XML y extraer su información. • Leer y extraer información de páginas Web. • Extracción y/o generación de datasets. • Verificar el contenido y la integridad de un dataset. • Analizar estadísticamente un dataset.

- Generar la predicción de datos usando regresión lineal.
- Predecir eventos utilizando Redes Bayesianas.
- Utilizar Redes Neuronales para la detección de patrones.
- Conocer y generar animaciones de imágenes en tiempo real usando redes convolucionales y GAN's.

9. Proyecto de asignatura

El objetivo del proyecto que planteé el docente que imparta esta asignatura, es demostrar el desarrollo y alcance de la(s) competencia(s) de la asignatura, considerando las siguientes fases:

Fundamentación: marco referencial (teórico, conceptual, contextual, legal) en el cual se fundamenta el proyecto de acuerdo con un diagnóstico realizado, mismo que permite a los estudiantes lograr la comprensión de la realidad o situación objeto de estudio para definir un proceso de intervención o hacer el diseño de un modelo.

Planeación: con base en el diagnóstico en esta fase se realiza el diseño del proyecto por parte de los estudiantes con asesoría del docente; implica planificar un proceso: de intervención empresarial, social o comunitario, el diseño de un modelo, entre otros, según el tipo de proyecto, las actividades a realizar los recursos requeridos y el cronograma de trabajo.

Ejecución: consiste en el desarrollo de la planeación del proyecto realizada por parte de los estudiantes con asesoría del docente, es decir en la intervención (social, empresarial), o construcción del modelo propuesto según el tipo de proyecto, es la fase de mayor duración que implica el desempeño de las competencias genéricas y específicas a desarrollar.

Evaluación: es la fase final que aplica un juicio de valor en el contexto laboral-profesional, social e investigativo, esta se debe realizar a través del reconocimiento de logros y aspectos a mejorar se estará promoviendo el concepto de “evaluación para la mejora continua”, la metacognición, el desarrollo del pensamiento crítico y reflexivo en los estudiantes.

10. Evaluación por competencias

La evaluación debe ser continua y permanente por lo que se debe considerar el desempeño en cada una de las actividades de aprendizaje, haciendo especial énfasis en:

- Participación en clase.
- Ejercicios realizados en clase.
- Información obtenida durante las búsquedas encomendadas.

- Lectura y análisis de textos.
- Autoevaluación, Coevaluación y evaluación de las actividades.

Revisión periódica del avance del proyecto (o proyectos) de la asignatura.

11. Fuentes de información

- El poder de los datos, Eloi Puertas, Oriol Pujol Vila, Santo Seguí y Jordi Vitriá, Primera Edición, RBA Contenidos Editoriales y Audiovisuales, España 2017, ISBN: 978-84-473-8940-7, Páginas: 140
- Intro to Python for Computer Science and Data Science, Paul Deitel and Harvey Deitel, First Edition, Person Education, USA 2020, ISBN-10: 0-13-540467-3, ISBN-13 978-0-13-540467-6, Páginas: 831
- Sistemas Distribuidos - Conceptos y Diseño, George Coulouris, Jean Dollimore y Tim Kindberg, Tercera Edición Addison Wesley Pearson Educación, Madrid España 2001, ISBN: 84-7829-049-4
- Practical Statistics for Data Scientists: 50+ Essential Concepts Using R and Python, Peter Bruce, Andrew Bruce, Peter Gedeck, Second Edition, O'Reilly Media, USA 2020, ISBN 13: 9781492072942
- Hands-On Machine Learning with Sci-kit Learn, Keras & TensorFlow, Aurélien Géron, Second Edition Seventh Release, O'Reilly Media, Sebastopol California USA, September 2019, ISBN: 978-1-492-03264-9, Páginas: 820
- Python Data Science Handbook, Jake VanderPlas, 2nd Edition, O'Reilly Media, Inc., USA 2022, ISBN 13: 9781098121211, páginas: 252
- Python for Data Analysis, Wes McKinney, 3rd Edition, O'Reilly Media, Inc., USA 2021, ISBN 13: 9781098104023, páginas: 361
- Python Deep Learning, Jordi Torres, Primera Edición en México, Alfaomega Grupo Editor, Ciudad de México abril 2020, ISBN: 978-607-538-613-3, Páginas: 384
- Introducción al aprendizaje por refuerzo profundo: Teoría y práctica en Python, Jordi Torres, Primera Edición Independently Published, Barcelona España abril 2021, ISBN: 979-8599775416, Páginas: 324
- Practical Natural Language Processing, Sowmya Vajjala, Bodhisattwa Majumder, Anuj Gupta, and Harshit Surana, First Edition, O'Reilly Media, Sebastopol California USA, June 2020, ISBN: 978-1-492-05405-4, Páginas: 455
- Cómo programar Internet & World Wide Web, Paul Deitel, Harbey Deitel y Abbey Deitel, Quinta Edición, Pearson Educación de México, Naucalpan de Juárez, Estado de México, septiembre 2013, ISBN: 978-607-32-2290-7