



1. Datos Generales de la asignatura

Nombre de la asignatura:	Introducción a la Ingeniería de Ciencia de Datos
Clave de la asignatura:	IAG-2501
SATCA¹:	3 – 3 – 6
Carrera:	Ingeniería Informática

2. Presentación

Caracterización de la asignatura
La asignatura "Introducción a la Ingeniería de Ciencia de Datos" proporciona una formación integral en el manejo de datos dentro del ámbito de la ingeniería, combinando fundamentos teóricos, habilidades técnicas y un enfoque crítico sobre su impacto en la sociedad. Su objetivo principal es que los alumnos comprendan el ciclo de vida de los datos, desde su adquisición hasta su análisis y visualización, utilizando herramientas modernas como Python, SQL, sistemas en la nube y frameworks de procesamiento masivo. Además de desarrollar competencias técnicas, los alumnos desarrollarán el pensamiento crítico y la toma de decisiones basada en datos, considerando aspectos éticos, de privacidad y de regulación en la gestión de la información. Con el trabajo en equipo, la resolución de problemas y la aplicación de metodologías ágiles, los alumnos se prepararán para enfrentar los desafíos tecnológicos actuales en entornos industriales y de investigación, promoviendo un uso responsable e innovador de la Ciencia de Datos en la ingeniería.
Intención didáctica
La asignatura "Introducción a la Ingeniería de Ciencia de Datos" tiene como objetivo proporcionar a los estudiantes una formación integral que combine el dominio de herramientas y técnicas para la gestión de datos con una visión crítica sobre su impacto en la ingeniería y la sociedad. Por medio de la práctica, los estudiantes aprenderán a adquirir, almacenar, procesar y analizar datos utilizando lenguajes de programación y plataformas modernas, preparando soluciones escalables para problemas reales.

¹ Sistema de Asignación y Transferencia de Créditos Académicos



Además, se fomentará la reflexión sobre la ética, la privacidad y la toma de decisiones basada en datos, abordando temas como la equidad en los modelos algorítmicos y el impacto social de la automatización. Mediante el desarrollo de proyectos, casos de estudio y actividades colaborativas, los estudiantes fortalecerán tanto su capacidad técnica como su pensamiento crítico, adquiriendo habilidades clave para la industria y la investigación en Ciencia de Datos.

En la unidad 1, los alumnos comprenderán el alcance y la importancia de la ciencia de datos dentro del ámbito de la ingeniería, proporcionándoles el contexto necesario para entender cómo la manipulación de datos impacta la toma de decisiones y la innovación tecnológica.

En la unidad 2, conocerán cómo obtener, almacenar y gestionar datos de manera eficiente y segura, preparando a los alumnos para manejar grandes volúmenes de información y garantizar su integridad, disponibilidad y confidencialidad.

En la unidad 3, se introduce a los alumnos técnicas modernas para la limpieza y transformación de datos a gran escala, dejando claro que el procesamiento eficiente de datos es clave para garantizar su utilidad en análisis y modelos predictivos.

En la unidad 4, los alumnos sabrán la importancia de interpretar y comunicar datos de manera efectiva, dándoles las habilidades clave para generar información valiosa y tomar decisiones informadas basadas en datos.

En la unidad 5, se abordan los aspectos prácticos y éticos de la implementación de sistemas de ciencia de datos en la industria. Los conocimientos en infraestructura y despliegue permiten a los ingenieros convertir modelos y análisis en soluciones escalables y reproducibles.



3. Participantes en el diseño y seguimiento curricular del programa

Lugar y fecha de elaboración o revisión	Participantes	Observaciones
TecNM Campus Iztapalapa III Febrero 2025.	Federico Mejía Cruz, Edeher Jim Bautista Olalde, Sandra Valeria Arenas Robles, Armando Montes Rivera, Sandra Magali García García, Néstor Josué Beltrán Espinoza, Brenda Erika Benítez Morales, Layonet Salvador Buenrostro Pérez, Sergio Alejandro Sánchez Rodríguez.	Reuniones presenciales con especialistas en la materia.

4. Competencia(s) a desarrollar

Competencia(s) específica(s) de la asignatura
Al finalizar el curso, el alumno será capaz de: • Comprender y aplicar los conceptos fundamentales de la Ciencia de Datos, incluyendo su ciclo de vida y su relación con otras disciplinas. • Utilizar herramientas de programación (Python con librerías como Pandas y NumPy) para la manipulación, limpieza y transformación de datos de manera eficiente. • Aplicar técnicas de estadística descriptiva para realizar la exploración inicial de conjuntos de datos, identificar patrones y resumir información relevante. • Implementar y evaluar modelos predictivos básicos (regresión lineal y algoritmos de clasificación fundamentales) utilizando herramientas de aprendizaje automático. • Crear visualizaciones efectivas utilizando librerías de Python para comunicar hallazgos y patrones identificados en los datos a diferentes audiencias. • Desarrollar habilidades de pensamiento crítico para interpretar resultados de análisis y modelos, así como para identificar posibles sesgos y limitaciones. • Comprender los aspectos éticos y las implicaciones sociales relacionadas con la recopilación, análisis y uso de datos.

5. Competencias previas



Para cursar la asignatura "**Introducción a la Ingeniería de Ciencia de Datos**", se recomienda que los alumnos cuenten con los siguientes conocimientos técnicos y habilidades generales, que les permitirán aprovechar al máximo los contenidos del curso.

Conocimientos Técnicos:

- **Fundamentos de programación:** Uso básico de lenguajes como Python, R o cualquier otro lenguaje de programación.
- **Matemáticas y estadística básica:** Conceptos fundamentales de álgebra lineal, cálculo, probabilidad y estadística.
- **Bases de datos:** Conocimientos elementales de SQL y modelos de bases de datos relacionales.
- **Manejo de herramientas digitales:** Uso de hojas de cálculo y software estadístico para manipulación y análisis de datos.

Habilidades Generales:

- **Pensamiento lógico y analítico:** Capacidad para estructurar problemas y diseñar soluciones basadas en datos.
- **Aprendizaje autónomo y autoorganización:** Disposición para explorar nuevas herramientas y metodologías en un campo en constante evolución.
- **Capacidad de resolución de problemas:** Aplicación de conocimientos matemáticos y computacionales en situaciones prácticas.
- **Interés por la aplicación de datos en problemas reales:** Actitud crítica y reflexiva sobre el impacto de la Ciencia de Datos en la ingeniería y la sociedad.

6. Temario

No.	Temas	Subtemas
1.	Fundamentos de la Ciencia de Datos y su Aplicación en Ingeniería	1.1. Conceptos básicos de la Ciencia de Datos 1.1.1 Definición y objetivos 1.1.2 Diferencias entre Ciencia de Datos, Ingeniería de Datos e Inteligencia Artificial 1.2. Ciclo de vida de los datos 1.2.1 Adquisición, almacenamiento, procesamiento y análisis 1.3. Aplicaciones de la Ciencia de Datos en la Ingeniería 1.3.1 Uso en sectores como manufactura, salud, finanzas y telecomunicaciones
2.	Adquisición, Almacenamiento y Gobernanza de Datos	2.1 Fuentes de datos y métodos de adquisición



		2.1.1 Datos estructurados y no estructurados 2.1.2 APIs, sensores IoT y bases de datos 2.2 Sistemas de almacenamiento de datos 2.2.1 Bases de datos relacionales y NoSQL 2.2.2 Data Lakes y Data Warehouses 2.3. Gobernanza y calidad de los datos 2.3.1 Seguridad, privacidad y ética en la gestión de datos 2.3.2 Normativas y estándares (GDPR, ISO 27001)
3.	Procesamiento y Transformación de Datos	3.1. Procesamiento de datos a gran escala 3.1.1 Conceptos de ETL (Extract, Transform, Load) 3.1.2 Frameworks como Apache Spark y Hadoop 3.2. Preprocesamiento y limpieza de datos 3.2.1 Detección y manejo de valores nulos y atípicos 3.2.2 Transformaciones y normalización 3.3 Automatización de flujos de datos 3.3.1 Herramientas de orquestación (Airflow, Prefect, Luigi)
4.	Análisis y Visualización de Datos	4.1. Introducción al análisis de datos 4.1.1 Estadística descriptiva e inferencial aplicada 4.1.2 Conceptos básicos de Machine Learning 4.2. Herramientas de análisis y visualización 4.2.1 Python (Pandas, NumPy, Matplotlib, Seaborn) 4.2.2 Power BI y Tableau 4.3. Narrativa con datos y comunicación efectiva 4.3.1 Storytelling con datos 4.3.2 Creación de reportes e informes ejecutivos
5.	Infraestructura y Despliegue en Ciencia de Datos	5.1. Computación en la nube para Ciencia de Datos



		5.1.1 AWS, Azure, Google Cloud 5.1.2 Servicios gestionados para almacenamiento y cómputo 5.2. Implementación de modelos en producción 5.2.1 Contenedores y microservicios (Docker, Kubernetes) 5.2.2 MLOps y CI/CD en Ciencia de Datos 5.3 Ética y responsabilidad en Ciencia de Datos 5.3.1 Impacto social de la IA y la automatización 5.3.2 Consideraciones de sesgo y equidad en los modelos
--	--	--

7. Actividades de aprendizaje de los temas

1. Fundamentos de la Ciencia de Datos y su Aplicación en Ingeniería	
Competencias	Actividades de aprendizaje
Específica(s): Comprender los conceptos fundamentales de la Ciencia de Datos y su impacto en la ingeniería, identificando su aplicación en la resolución de problemas del mundo real. Genéricas: • Comunicación oral y escrita. • Habilidad para buscar y analizar información proveniente de fuentes diversas. • Capacidad de aplicar los conocimientos en la práctica.	• Análisis de artículos sobre el impacto de la Ciencia de Datos en la industria y la ingeniería. • Elaboración de un esquema que represente el ciclo de vida de los datos en diferentes aplicaciones. • Identificación de ejemplos reales donde la Ciencia de Datos ha generado innovación o eficiencia en el sector industrial. • Exploración inicial de un conjunto de datos reales para identificar patrones y problemas comunes. • Investigación sobre avances recientes en Ciencia de Datos y su impacto en la toma de decisiones.
2. Adquisición, Almacenamiento y Gobernanza de Datos	
Competencias	Actividades de aprendizaje
Específica(s): Aplicar técnicas de adquisición y almacenamiento de datos en bases de datos estructuradas y no estructuradas, considerando principios de gobernanza, privacidad y seguridad de la información. Genéricas: • Comunicación oral y escrita. • Habilidad para buscar y analizar información proveniente de fuentes diversas. • Capacidad de aplicar los conocimientos en la práctica.	• Identificación y clasificación de fuentes de datos estructurados y no estructurados. • Ejercicios básicos de consultas, inserción y manipulación de datos. • Diseño de un modelo de datos optimizado para un caso práctico. • Discusión sobre la gobernanza de datos y su regulación en distintos países. • Uso de herramientas como Postman o Python para extraer información desde servicios web.
3. Procesamiento y Transformación de Datos	
Competencias	Actividades de aprendizaje
Específica(s): Desarrollar habilidades para limpiar, transformar y preparar datos mediante herramientas y lenguajes de programación, optimizando su calidad para su posterior análisis.	• Identificación y corrección de valores nulos, duplicados y atípicos en un dataset.

Genéricas: • Comunicación oral y escrita. • Habilidad para buscar y analizar información proveniente de fuentes diversas. • Capacidad de aplicar los conocimientos en la práctica.	• Implementación de un flujo de datos utilizando Apache Spark o Pandas en Python. • Creación de un script en Python para extraer, transformar y cargar datos en una base de datos. • Aplicación de diferentes métodos de transformación de datos y análisis de su impacto. • Aplicación de todas las técnicas aprendidas para preparar un dataset listo para su análisis.
4. Análisis y Visualización de Datos	
Competencias	Actividades de aprendizaje
Específica(s): Analizar e interpretar datos utilizando herramientas de análisis estadístico y técnicas de visualización, generando reportes y dashboards que faciliten la toma de decisiones basada en datos. Genéricas: • Comunicación oral y escrita. • Habilidad para buscar y analizar información proveniente de fuentes diversas. • Capacidad de aplicar los conocimientos en la práctica.	• Uso de Pandas y NumPy para obtener estadísticas descriptivas de un conjunto de datos. • Creación de visualizaciones con Matplotlib y Seaborn para interpretar tendencias. • Presentación de un informe donde se expliquen los hallazgos obtenidos a partir de un dataset. • Uso de Power BI o Tableau para diseñar un dashboard interactivo. • Análisis crítico de visualizaciones de datos para identificar posibles sesgos o errores.
5. Infraestructura y Despliegue en Ciencia de Datos	
Competencias	Actividades de aprendizaje
Específica(s): Implementar soluciones para el procesamiento y despliegue de modelos de datos en entornos productivos, utilizando tecnologías de computación en la nube y contenedores para su escalabilidad y eficiencia. Genéricas: • Comunicación oral y escrita. • Habilidad para buscar y analizar información proveniente de fuentes diversas.	• Creación de una cuenta en AWS, Azure o Google Cloud para familiarizarse con los servicios disponibles. • Creación de un contenedor con un entorno de Ciencia de Datos preconfigurado. • Implementación básica de un modelo en producción usando Flask o FastAPI. • Evaluación de casos donde el uso de datos ha generado impactos negativos.



• Capacidad de aplicar los conocimientos en la práctica.	• Desarrollo de un flujo completo desde la adquisición hasta la visualización y despliegue de datos.
--	--

8. Práctica(s)

Se hace énfasis en la realización de prácticas de campo con datos reales, permitiendo que el alumno experimente de primera mano el proceso de adquisición, limpieza, análisis y visualización de datos. A través de estas actividades, se fomenta una comprensión profunda sobre la aplicabilidad de la Ciencia de Datos en diversos ámbitos, como la industria, la salud, la economía y la tecnología.

Además, se busca que los alumnos desarrollen una visión crítica sobre la calidad y veracidad de la información, fortaleciendo sus habilidades para la toma de decisiones basada en datos en un mundo cada vez más impulsado por la información.



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-profesión, social e investigativo, ésta 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

El docente que imparta esta asignatura posee toda una gama de herramientas que puede utilizar para la adecuada evaluación de sus alumnos.

Estrategias de Evaluación:

- Evaluación Diagnóstica
- Evaluación Formativa
 - Observación en las sesiones prácticas
 - Análisis de tareas y ejercicios
 - Participación en clase y discusiones
 - Evaluación entre pares (coevaluación)
- Evaluación Sumativa:
 - Exámenes teórico-prácticos
 - Proyectos prácticos
 - Informes técnicos
 - Presentaciones orales

Instrumentos de Evaluación:

- Listas de cotejo
- Rúbricas o matrices de valoración
- Guías de observación
- Portafolios de evidencias

Autoevaluación

11. Fuentes de información

Impresas:

1. Zumel, N., & Mount, J. (2014). Practical Data Science with R. Manning Publications.
2. McKinney, W. (2017). Python for Data Analysis: Data Wrangling with Pandas, NumPy, and IPython. O'Reilly Media.
3. Bruce, P., & Bruce, A. (2017). Practical Statistics for Data Scientists: 50 Essential Concepts. O'Reilly Media.
4. Wilke, C. O. (2019). Fundamentals of Data Visualization: A Primer on Making Informative and Compelling Figures. O'Reilly Media.
5. Provost, F., & Fawcett, T. (2013). Data Science for Business: What You Need to Know About Data Mining and Data-Analytic Thinking. O'Reilly Media.

Electrónicas:

1. Documentación oficial de las librerías de Python (Pandas, NumPy, Scikit-learn, Matplotlib, Seaborn).
2. Cursos en línea y tutoriales de plataformas como Coursera, edX, DataCamp, Udacity.
3. Blogs y comunidades de Ciencia de Datos (Towards Data Science, Kaggle).
4. Artículos de investigación y preprints en arXiv y otras plataformas.