



1. Datos Generales de la asignatura

Nombre de la asignatura:	Adquisición de datos
Clave de la asignatura:	CDD-2401
SATCA¹:	2-3-5
Carrera:	Ingeniería en Ciencia de Datos

2. Presentación

Caracterización de la asignatura
<p>La asignatura ofrece a los estudiantes una comprensión sólida de las técnicas y estrategias utilizadas para recopilar datos de diversas fuentes, en el contexto de la Ciencia de Datos. A lo largo del curso, los estudiantes adquieren habilidades prácticas en la adquisición, limpieza y preprocesamiento de datos, lo que les permite preparar datos de alta calidad para su posterior análisis.</p> <p>Para cursar esta asignatura se requiere que los estudiantes tengan conocimientos previos en estadística y programación, particularmente en lenguajes como Python o R.</p>
Intención didáctica
<p>La intención didáctica de esta asignatura es proporcionar a los estudiantes una comprensión sólida y habilidades prácticas en la obtención de datos, ética en la recopilación de información y preprocesamiento de datos para su uso en proyectos de Ciencia de Datos. A lo largo de las unidades del curso, los estudiantes adquirirán conocimientos fundamentales y técnicas avanzadas para adquirir datos de diversas fuentes y prepararlos para su análisis. El objetivo es formar a los estudiantes para que sean capaces de recopilar y limpiar datos de alta calidad que sean fundamentales en el éxito de cualquier proyecto de Ciencia de Datos.</p> <p>La asignatura está organizada en cinco unidades que abarcan desde los fundamentos de la obtención de datos hasta la aplicación de técnicas especializadas en proyectos prácticos. A lo largo del curso, los estudiantes adquirirán las siguientes competencias:</p> <p>Comprensión de los Fundamentos de Obtención de Datos: Los estudiantes comprenderán los conceptos fundamentales de la obtención de datos, incluyendo fuentes de datos, métodos de recopilación y la importancia de datos limpios y precisos en la ciencia de datos.</p> <p>Dominio de técnicas de limpieza y preprocesamiento: Los estudiantes adquirirán habilidades prácticas en la limpieza y preprocesamiento de datos, lo que les permitirá abordar problemas comunes como valores faltantes, duplicados y ruido en los datos.</p>

¹ Sistema de Asignación y Transferencia de Créditos Académicos



Conciencia ética y legal en la obtención de datos: Se fomentará una comprensión sólida de las consideraciones éticas y legales en la recopilación de datos, incluyendo la privacidad y la seguridad de la información.

Recopilación de Datos Especializada: Los estudiantes explorarán técnicas avanzadas de recopilación de datos, adaptadas a áreas específicas de interés, como la ecología, la salud o la tecnología.

Aplicación Práctica en Proyectos: La asignatura culminará en la realización de proyectos prácticos donde los estudiantes aplicarán las técnicas aprendidas en situaciones del mundo real, desde la obtención de datos hasta el análisis y la presentación de resultados.

A través de una combinación de teoría, ejercicios prácticos y proyectos, los estudiantes desarrollarán habilidades esenciales para la obtención y preparación efectiva de datos, asegurando que estén equipados para abordar desafíos reales en el campo de la ciencia de datos.

3. Participantes en el diseño y seguimiento curricular del programa

Lugar y fecha de elaboración o revisión	Participantes	Observaciones
Instituto Tecnológico Superior de Alvarado del 21 al 23 agosto de 2023.	Representante del Instituto Tecnológico Superior de Alvarado.	Propuesta inicial.
Tecnológico Nacional de México 30 octubre 2023	Representante del Instituto Tecnológico de: Querétaro y del Instituto Tecnológico Superior de Alvarado.	Presentación de la propuesta de la carrera de Ingeniería en Ciencia de Datos.
Instituto Tecnológico de Querétaro Campus Norte del 19 al 22 de marzo 2024.	Representantes de los Institutos Tecnológicos de: Morelia, Puebla, Querétaro, Tehuacán. Instituto Tecnológico Superior de Alvarado. CENIDET. Representante de Ciencias Básica de los Institutos de: Celaya, Morelia y CIIDET.	Diseño y/o desarrollo curricular de la carrera de Ingeniería en Ciencia de Datos.
Tecnológico Nacional de México del 22 al 24 de abril del 2024	Representante del Instituto Tecnológico de Querétaro e Instituto Tecnológico Superior de Alvarado.	Contraste y ajuste de las asignaturas de Ingeniería en Ciencia de Datos con respecto a las de Ing. en Inteligencia Artificial, Ing. en Desarrollo WEB e Ing. en Ciberseguridad



Tecnológico Nacional de México del 27 al 31 de mayo del 2024.	Representantes de los Institutos Tecnológicos de: Morelia, Querétaro. Instituto Tecnológico Superior de Alvarado. CENIDET.	Consolidación curricular de la carrera de Ingeniería Ciencia de Datos
---	--	---

4. Competencia(s) a desarrollar

Competencia(s) específica(s) de la asignatura
Desarrolla estrategias efectivas para la obtención de datos de diversas fuentes, aplicando técnicas avanzadas de limpieza y preprocesamiento de datos para garantizar la calidad y la utilidad de los datos en proyectos de Ciencia de Datos., abordando las consideraciones éticas y legales en la obtención de datos, promoviendo prácticas de privacidad y cumplimiento normativo.

5. Competencias previas

<p>Aplica algoritmos y lenguajes de programación para la implementación en el diseño de soluciones problemáticas reales de la sociedad.</p> <p>Habilidad de pensamiento lógico y resolución de problemas.</p> <p>Analiza los requerimientos, manipula bases de datos para brindar soluciones al tratamiento de información basándose en modelos y estándares; de igual modo, implementa bases de datos para apoyar la toma de decisiones considerando las reglas de negocio.</p>
--

6. Temario

No.	Temas	Subtemas
1	Fundamentos de obtención de datos.	1.1. Introducción a la obtención de datos. 1.1.1. Definición y conceptos básicos de obtención de datos. 1.1.2. Importancia de la obtención de datos en Ciencia de Datos. 1.2. Fuentes de datos. 1.2.1. Diversidad de fuentes de datos: Web, APIs, bases de datos, archivos, sensores, etc. 1.2.2. Desafíos en la obtención de datos.
2	Técnicas de recopilación de datos.	2.1. Web Scraping y APIs. 2.1.1. Web Scraping: recopilación de datos de sitios web. 2.1.2. Uso de APIs: acceso a datos estructurados. 2.2. Obtención de datos de diversas fuentes. 2.2.1. Obtención de datos de archivos locales y remotos. 2.2.2. Adquisición de datos de fuentes en tiempo real.



3	Limpieza y preprocesamiento de datos.	<ul style="list-style-type: none">3.1. Calidad y preparación de datos.<ul style="list-style-type: none">3.1.1. Identificación y manejo de valores faltantes.3.1.2. Detección y tratamiento de valores atípicos y ruido.3.1.3. Normalización y estandarización de datos.3.1.4. Discretización de datos y jerarquías conceptuales.3.1.5. Reducción de datos.3.2. Integración y validación de datos.<ul style="list-style-type: none">3.2.1. Integración de datos de múltiples fuentes.3.2.2. Validación de datos para garantizar integridad y precisión.
4	Ética y legalidad en la obtención de datos.	<ul style="list-style-type: none">4.1. Ética y cumplimiento normativo.<ul style="list-style-type: none">4.1.1. Consideraciones éticas en la recopilación de datos.4.1.2. Cumplimiento normativo y regulaciones de privacidad.4.2. Responsabilidad y protección de datos.<ul style="list-style-type: none">4.2.1. Responsabilidad en el uso de datos.4.2.2. Protección de datos personales y confidenciales.
5	Recopilación de datos especializada y proyectos prácticos.	<ul style="list-style-type: none">5.1. Obtención de datos especializada.<ul style="list-style-type: none">5.1.1. Obtención de datos de redes sociales (por ejemplo, Twitter, Facebook).5.1.2. Adquisición de datos geoespaciales y de mapas.5.1.3. Uso de APIs avanzadas para datos especializados.5.2. Proyectos prácticos de obtención de datos.<ul style="list-style-type: none">5.2.1. Desarrollo de proyectos prácticos que involucren la obtención de datos de fuentes diversas.5.2.2. Aplicación de técnicas de limpieza y preprocesamiento en datos reales.5.2.3. Presentación de resultados y análisis de datos obtenidos.



7. Actividades de aprendizaje de los temas

1. Fundamentos de obtención de datos.	
Competencias	Actividades de aprendizaje
<p><i>Específica(s):</i> Comprende las diferentes fuentes y estrategias de obtención de datos y la importancia de la obtención de datos en el contexto de la Ciencia de Datos.</p> <p><i>Genérica(s):</i></p> <ul style="list-style-type: none"> ● Comprensión conceptual. ● Identificación de problemas. ● Capacidad para trabajar en equipo ● Habilidad para buscar y analizar información proveniente de fuentes diversas. ● Resolución de problemas básicos. ● Pensamiento lógico. 	<ul style="list-style-type: none"> ● Investigación: los estudiantes investigarán y documentarán diferentes fuentes de datos utilizadas en ciencia de datos, como bases de datos, APIs, fuentes web, sensores, archivos, etc. Deberán elaborar una lista de ejemplos y describir las características de cada fuente. ● Casos de estudio: los estudiantes analizarán casos de estudio reales de obtención de datos en proyectos de ciencia de datos. Esto puede incluir ejemplos de aplicaciones en áreas como el análisis financiero, la salud, la inteligencia empresarial, etc. ● Discusión en plenaria (debate): se organizará una discusión en grupo donde los estudiantes compartirán sus hallazgos sobre las fuentes de datos y los casos de uso analizados. Deberán discutir la importancia de la obtención de datos en estos contextos. ● Ejercicio: los estudiantes recibirán un escenario hipotético y deberán seleccionar la fuente de datos más adecuada para ese escenario. Esto fomentará la identificación de problemas y la toma de decisiones informadas. ● Presentación: cada estudiante preparará una breve presentación sobre la importancia de la obtención de datos en ciencia de Datos. Deberán destacar ejemplos específicos que ilustren su importancia. ● Resumen: los estudiantes resumirán las estrategias de obtención de datos discutidas, incluyendo los desafíos asociados.



2. Técnicas de recopilación de datos	
Competencias	Actividades de aprendizaje
<p><i>Específica(s):</i></p> <ul style="list-style-type: none"> • Domina técnicas de recopilación y extracción de datos. • Selecciona la técnica de recopilación adecuada según el tipo base de datos. <p><i>Genérica(s):</i></p> <ul style="list-style-type: none"> • Habilidad para buscar y analizar información proveniente de fuentes diversas. • Pensamiento lógico. • Resolución de problemas básicos. • Capacidad para trabajar en equipo. 	<ul style="list-style-type: none"> • Práctica de web scraping: los estudiantes realizarán ejercicios prácticos de web scraping para recopilar datos de sitios web específicos. Utilizarán herramientas y bibliotecas relevantes (por ejemplo, BeautifulSoup en Python) para extraer información estructurada de páginas web. Compartirán sus experiencias y desafíos con el grupo. • Caso de estudio: los estudiantes seleccionarán una API pública y desarrollarán un proyecto que involucre la obtención de datos a través de esa API. Aprenderán a autenticarse, enviar solicitudes y procesar respuestas de la API. Presentarán sus proyectos y los datos obtenidos en clase. • Práctica de extracción de datos: Los estudiantes trabajarán con una base de datos de muestra y realizarán consultas para extraer datos específicos. Presentarán los resultados de sus consultas y discutirán los desafíos que enfrentaron. • Práctica de obtención de datos en tiempo real: los estudiantes explorarán cómo obtener datos en tiempo real de fuentes como sensores IoT o transmisiones en línea. Desarrollarán aplicaciones simples que utilicen datos en tiempo real. Presentarán sus aplicaciones y compartirán su comprensión de los datos en tiempo real. • Cuadro comparativo: realizarán un ejercicio de comparación de técnicas de recopilación de datos. Se les dará un escenario y deberán determinar la técnica más adecuada (web scraping, API, base de datos) para obtener los datos requeridos. Discutirán sus elecciones y argumentarán sus decisiones.



3. Limpieza y preprocesamiento de datos	
Competencias	Actividades de aprendizaje
<p><i>Específica(s):</i> Maneja las diferentes técnicas de preprocesamiento de datos, para obtener resultados de calidad.</p> <p><i>Genérica(s):</i></p> <ul style="list-style-type: none"> • Habilidad para buscar y analizar información proveniente de fuentes diversas. • Resolución de problemas básicos. • Pensamiento lógico. • Capacidad para trabajar en equipo. 	<ul style="list-style-type: none"> • Ejercicio: Los estudiantes recibirán conjuntos de datos que contienen valores atípicos y trabajarán en la identificación de estos valores. Utilizarán herramientas y técnicas estadísticas para detectar valores extremos y atípicos. Presentarán informes que destaquen los valores atípicos identificados y las estrategias para manejarlos. • Práctica: Los estudiantes trabajarán con conjuntos de datos que requieren normalización y estandarización. Aplicarán técnicas para escalar datos en rangos uniformes y estandarizarlos para comparaciones efectivas. Presentarán los datos normalizados y estandarizados junto con análisis comparativos. • Práctica: Los estudiantes recibirán datos de múltiples fuentes que deben integrarse en un solo conjunto de datos. Identificarán conflictos de datos y aplicarán técnicas de resolución. Presentarán el conjunto de datos integrado y discutirán los desafíos en el proceso de integración. • Caso de estudio: Los estudiantes aplicarán las técnicas de limpieza y preprocesamiento de datos aprendidas en proyectos prácticos de obtención de datos. Identificarán y resolverán problemas específicos de calidad de datos en conjuntos reales. Presentarán los resultados de la limpieza y cómo afectaron los análisis posteriores. • Casos de estudio: Los estudiantes analizarán cómo la limpieza y preprocesamiento de datos afectan los resultados y conclusiones en proyectos de Ciencia de Datos. Discutirán ejemplos en los que la falta de limpieza de datos puede llevar a interpretaciones erróneas. Presentarán casos de estudio y conclusiones sobre la importancia de la limpieza de datos.



4. Ética y legalidad en la obtención de datos	
Competencias	Actividades de aprendizaje
<p><i>Específica(s):</i> Aplica principios éticos en la obtención de datos, cumpliendo con regulaciones de privacidad y normativas relacionadas con el proceso de recopilación de datos.</p> <p><i>Genéricas:</i></p> <ul style="list-style-type: none"> ● Pensamiento ético. ● Comunicación efectiva. ● Capacidad para trabajar en equipo. ● Habilidad para buscar y analizar información proveniente de fuentes diversas. 	<ul style="list-style-type: none"> ● Caso de estudio: Los estudiantes analizarán casos éticos relacionados con la obtención de datos en proyectos de Ciencia de Datos. Identificarán dilemas éticos y tomarán decisiones basadas en principios éticos. Presentarán sus análisis y discutirán las implicaciones éticas de cada caso. ● Investigación: Los estudiantes investigarán regulaciones de privacidad y normativas relacionadas con la obtención de datos, como el Reglamento General de Protección de Datos (GDPR) de la Unión Europea. Comprenderán las responsabilidades legales y las sanciones por incumplimiento. Prepararán un informe que destaque los aspectos clave de las regulaciones y cómo se aplican en la obtención de datos. ● Ejercicio: Los estudiantes trabajarán en grupos para desarrollar códigos de ética para la obtención de datos en proyectos de Ciencia de Datos. Considerarán aspectos como la privacidad, la transparencia y el consentimiento. Presentarán sus códigos de ética y discutirán su aplicabilidad en situaciones reales. ● Presentación: Los estudiantes realizarán proyectos prácticos de obtención y análisis de datos. Deberán presentar sus resultados de manera ética, incluyendo la divulgación de fuentes de datos y prácticas de obtención. Recibirán retroalimentación sobre la presentación ética de sus proyectos. ● Debate: Los estudiantes participarán en discusiones sobre el impacto social de la obtención de datos y el uso de datos personales. Analizarán temas como la discriminación algorítmica y la responsabilidad en el uso de datos. Compartirán sus opiniones y soluciones propuestas.



5. Recopilación de datos especializada y proyectos prácticos	
Competencias	Actividades de aprendizaje
<p><i>Específica(s):</i></p> <ul style="list-style-type: none"> ● Aplica técnicas avanzadas de obtención de datos especializadas. ● Diseña y ejecuta proyectos de obtención de datos. <p><i>Genéricas:</i></p> <ul style="list-style-type: none"> ● Habilidad para buscar y analizar información proveniente de fuentes diversas. ● La comprensión y manipulación de ideas y pensamientos. ● Metodologías para solución de problemas, organización del tiempo y para el aprendizaje. ● Habilidad en el manejo de equipo de cómputo. ● Capacidad para trabajar en equipo. ● Capacidad de aplicar los conocimientos en la práctica. 	<ul style="list-style-type: none"> ● Práctica: Los estudiantes explorarán técnicas de obtención de datos especializadas, como la obtención de datos de redes sociales (por ejemplo, Twitter o Facebook). Realizarán ejercicios prácticos de obtención de datos en estas áreas y aprenderán a utilizar API específicas. Presentarán los datos obtenidos y analizarán su relevancia. ● Práctica: Los estudiantes trabajarán con datos geoespaciales y aprenderán a obtener información de mapas, sistemas de información geográfica (SIG) o servicios de geolocalización. Realizarán trabajos que involucran la obtención de datos geoespaciales y su análisis. Presentarán sus trabajos y discutirán las aplicaciones de datos geoespaciales. ● Proyecto: Los estudiantes formarán equipos para diseñar y ejecutar proyectos de obtención de datos prácticos. Seleccionarán un tema y definirán objetivos claros. Obtendrán datos, los limpiarán y realizarán análisis preliminares. Presentarán los resultados de sus proyectos y discutirán los desafíos enfrentados durante el proceso. ● Ejercicio: Los estudiantes revisarán proyectos anteriores y buscarán formas de optimizar las técnicas de obtención de datos utilizadas. Identificarán mejoras en eficiencia y calidad de los datos. Presentarán sus hallazgos y soluciones propuestas. ● Presentación: Los estudiantes presentarán sus proyectos prácticos a la clase, compartiendo los datos obtenidos, los análisis realizados y las lecciones aprendidas. Participarán en discusiones y recibirán retroalimentación de sus compañeros y el profesor.



8. Práctica(s)

- Investigar y seleccionar una fuente de datos específica relacionada con un tema de interés (por ejemplo, datos gubernamentales, datos de redes sociales, datos de sensores, etc.).
- Identificar la estructura y el formato de los datos disponibles en la fuente seleccionada.
- Presentar un informe que describa la fuente de datos, su estructura y su relevancia para un posible proyecto de Ciencia de Datos.
- Diseñar un plan detallado para recopilar datos de la fuente seleccionada.
- Definir objetivos claros de recopilación, métodos de obtención de datos y las herramientas necesarias.
- Presentarán el plan de recopilación de datos, incluyendo consideraciones éticas y legales.
- Implementar el plan de recopilación de datos diseñado en otra práctica realizada anteriormente.
- Limpiar datos.
- Abordar problemas comunes de calidad de datos, como valores faltantes o duplicados.
- Identificar patrones o tendencias iniciales en los datos y presentar los hallazgos.

9. Proyecto de asignatura

El objetivo del proyecto que planteé el docente que imparta esta asignatura, es demostrar el desarrollo y alcance del(los) logro(s) formativo(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 los saberes, habilidades y destrezas 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”, el desarrollo del pensamiento crítico y reflexivo en los estudiantes.



10. Evaluación por competencias

La evaluación debe hacerse diagnóstica, formativa y sumativa. De igual manera, para fortalecer la parte actitudinal, se recomienda guiar al estudiante hacia la introspección para utilizar la autoevaluación y la coevaluación.

En el caso de las actividades de aprendizaje se sugiere el uso de estrategias metacognitivas como: mapas mentales, mapas conceptuales, reportes de prácticas, exposiciones en clase, ensayos, resúmenes, observación y cuestionarios, cuadros comparativos, informes.

Mientras que para verificar el nivel del logro de las competencias del estudiante se recomienda utilizar: el portafolio de evidencias, listas de cotejo, rúbricas, matrices de valoración, exámenes, guías de observación, además de estrategias en las que se logren las competencias blandas.

11. Fuentes de Información

1. Bart Baesens, S. v. (2018). Practical Web Scraping for Data Science: Best Practices and Examples with Python. Apress.
2. Castillo, J. N. (2020). Introducción a la ciencia de datos en R Un enfoque práctico. Editorial Universidad Distrital Francisco José de Caldas.
3. Creswell, J. W. (2013). Research Design: Qualitative, Quantitative, and Mixed Methods Approaches. Sage Pubns.
4. Davenport, T. H. (2014). Big Data at Work: Dispelling the Myths, Uncovering the Opportunities. Harvard Business Review Press.
5. Davis, K. (2012). Ethics of Big Data: Balancing Risk and Innovation. O'Reilly Media.
6. Gagolewski, M. (2017). Minimalist Data Wrangling with Python. O'Reilly Media.
7. Jarmul, K. J. (2017). Python Web Scraping - Second Edition: Hands-on data scraping and crawling using PyQT, Selenium, HTML and Python. Packt Publishing.
8. Lohr, S. L. (2021). Sampling: Design and Analysis (Chapman & Hall/CRC Texts in Statistical Science). Chapman and Hall/CRC.
9. Matthew A. Russell, M. K. (2019). Mining the Social Web, 3rd Edition. O'Reilly Media, Inc.
10. McKinney, W. (2022). Python for Data Analysis. O'Reilly Media.
11. Merriam, S. B. (2019). Qualitative Research: A Guide to Design and Implementation, 4th Edition. Gildan Media Corporation.
12. Mitchell, R. (2018). Web Scraping with Python. O'Reilly Media, Inc.
13. Olsen, W. K. (2011). Data Collection: Key Debates and Methods in Social Research. SAGE Publications Ltd.
14. Poynter, R. (2010). The Handbook of Online and Social Media Research: Tools and Techniques for Market Researchers. Wiley.
15. Robert M. Groves, F. J. (2009). Survey Methodology. WILEY.
16. Sapsford, R. (1996). Data Collection and Analysis (Published in association with The Open University). SAGE Publications Ltd