1. Datos Generales de la asignatura

Nombre de la asignatura:	Big data
Clave de la asignatura:	CDD-2404
SATCA[1]:	2-3-5
Carrera:	Ingeniería en Ciencia de Datos

2. Presentación

Caracterización de la asignatura

La asignatura de Big Data representa un componente fundamental en la formación de un Ingeniero en Ciencia de Datos, ya que aborda de manera profunda y especializada el manejo, análisis y aprovechamiento de conjuntos de datos masivos y complejos. Esta asignatura se centra en proporcionar al estudiante las herramientas, técnicas y metodologías necesarias para enfrentar los desafíos específicos asociados con el procesamiento y la extracción de conocimiento a partir de grandes volúmenes de datos, que van más allá de lo que las técnicas tradicionales de análisis de datos pueden abordar.

En primer lugar, la asignatura de Big Data introduce al estudiante en los conceptos fundamentales de la infraestructura tecnológica necesaria para gestionar eficientemente grandes cantidades de datos, incluyendo sistemas de almacenamiento distribuido, procesamiento en paralelo y herramientas de procesamiento de datos en tiempo real. Asimismo, se profundiza en el uso de tecnologías específicas como Hadoop, Spark y sistemas de bases de datos NoSQL, que son esenciales para manejar eficazmente la variedad, velocidad y volumen de los datos en entornos de Big Data.

Además de la infraestructura tecnológica, la asignatura de Big Data capacita al Ingeniero en Ciencia de Datos en técnicas avanzadas de análisis y minería de datos diseñadas específicamente para enfrentar los retos inherentes a conjuntos de datos masivos. Se exploran algoritmos de aprendizaje automático escalables, técnicas de procesamiento de lenguaje natural, entre otros, con el objetivo de extraer información valiosa, identificar patrones y realizar predicciones precisas a partir de datos de gran escala.

[1] Sistema de Asignación y Transferencia de Créditos Académicos

La importancia de esta asignatura para el Ingeniero en Ciencia de Datos radica en su relevancia directa para la resolución de problemas del mundo real en una amplia gama de campos, desde la industria y la investigación hasta la salud y el gobierno. En la era actual, donde la cantidad de datos generados crece exponencialmente cada día, la capacidad para comprender, analizar y extraer conocimiento significativo de estos datos se ha convertido en un activo invaluable para las organizaciones y la sociedad en general. Por lo tanto, la capacitación en Big Data permite al Ingeniero en Ciencia de Datos estar preparado para enfrentar los desafíos y aprovechar las oportunidades que surgen en este contexto de datos masivos, permitiéndoles ofrecer soluciones innovadoras y estratégicas que impulsen el progreso y la toma de decisiones informadas.

Intención didáctica

El temario está organizado en seis unidades de aprendizaje, cada una diseñada para proporcionar al estudiante los conocimientos, habilidades y herramientas necesarias para abordar diferentes aspectos del procesamiento y análisis de grandes volúmenes de datos.
En la primera unidad tiene como objetivo familiarizar al estudiante con los conceptos fundamentales del procesamiento de Big Data, incluyendo los desafíos y oportunidades asociados. Además, se exploran casos de uso y aplicaciones prácticas para comprender la relevancia de esta área en diversos contextos.

En la segunda unidad se profundiza en las técnicas y metodologías para el procesamiento eficiente de grandes volúmenes de datos en entornos distribuidos. Se estudian en detalle los sistemas de almacenamiento distribuido, el procesamiento en paralelo y las estrategias de escalabilidad horizontal. El objetivo es capacitar al estudiante en el diseño e implementación de soluciones escalables para el procesamiento de datos a gran escala.

En la tercera unidad se proporciona al estudiante las habilidades necesarias para optimizar el rendimiento y la eficiencia de los procesos de análisis de datos en entornos de Big Data. Se exploran técnicas de optimización de consultas, diseño de modelos de datos eficientes y estrategias para mejorar el rendimiento de algoritmos de análisis de datos a gran escala.

En la cuarta unidad, se aborda el estudio de algoritmos de aprendizaje automático supervisado diseñados para trabajar con conjuntos de datos masivos. Se exploran técnicas de clasificación y regresión, así como estrategias para entrenar y evaluar modelos de aprendizaje supervisado en entornos distribuidos. El objetivo es capacitar al estudiante en la aplicación de técnicas de aprendizaje enfocado a problemas de Big Data.

Para la quinta unidad tenemos el estudio de algoritmos de aprendizaje automático no supervisado adaptados para el análisis de grandes volúmenes de datos. Se exploran técnicas de clustering, reducción de dimensionalidad y detección de anomalías, así como su aplicación en la extracción de conocimiento a partir de datos no etiquetados a gran escala.

Y finalmente en la sexta unidad el propósito es proporcionar al estudiante las habilidades necesarias para el procesamiento y análisis de grandes cantidades de datos de texto. Se exploran técnicas de preprocesamiento de texto, extracción de características, análisis de sentimientos y clasificación de texto a gran escala. El objetivo es capacitar al estudiante en la aplicación de técnicas de minería de texto para extraer información significativa y conocimiento útil de grandes colecciones de datos de texto.

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
● Implementa infraestructuras tecnológicas para el procesamiento eficiente de grandes volúmenes de datos. ● Aplica técnicas avanzadas de procesamiento y análisis de datos a gran escala. ● Optimiza procesos de modelado de datos, tanto en contextos supervisados como no supervisados. ● Aplica algoritmos de aprendizaje supervisado y no supervisado a gran escala para la identificación de patrones y la realización de predicciones precisas. ● Utiliza técnicas de minería de texto a gran escala para extraer información relevante de documentos y fuentes de datos textuales. ● Interpreta resultados obtenidos y los aplica soportando la toma de decisiones, impulsando la innovación y el desarrollo en diversos ámbitos profesionales.

5. Competencias previas

- Programación: conocimiento básico de al menos un lenguaje de programación como Python, Java o R, que permita la implementación de algoritmos y el manejo de datos.
- Bases de Datos: comprender los principios de diseño y manipulación de bases de datos relacionales, así como nociones básicas de SQL para consultas de datos.
- Estadística y Probabilidad: entender los conceptos estadísticos básicos, como la distribución de datos, la probabilidad y la inferencia estadística, que son fundamentales para el análisis de datos.
- Álgebra Lineal: conocimientos básicos de álgebra lineal son útiles para comprender ciertos algoritmos y técnicas utilizadas en el procesamiento de datos y el aprendizaje automático.
- Fundamentos de Ciencia de Datos: familiaridad con conceptos básicos de ciencia de datos, como limpieza de datos, visualización de datos, y técnicas de preprocesamiento.
- Fundamentos de Aprendizaje Máquina: conocimientos introductorios sobre algoritmos de aprendizaje automático, como regresión, clasificación y agrupamiento, así como su implementación práctica.

6. Temario

No.	Temas	Subtemas
1	Introducción a Big Data.	1.1 Introducción a análisis de Big Data. 1.2 Conceptos básicos a tecnologías para el procesamiento con Big Data. 1.3 Fundamentos de programación. 1.4 Datos distribuidos resilientes.
2	Procesamiento a gran escala.	2.1. Frameworks de procesamiento a gran escala. 2.2. Álgebra lineal para datos a gran escala. 2.3. Sistemas distribuidos de almacenamiento de archivos.
3	Optimización y modelado de datos.	3.1. Introducción a modelado. 3.1.1. Numérico. 3.1.2. Probabilístico. 3.1.3. Bayesiano. 3.2. Introducción a problemas de optimización. 3.3. Descenso de gradiente estocástico por lotes. 3.4. Método de Newton. 3.5. Métodos de cadenas de Markov Monte Carlo.
4	Aprendizaje supervisado a gran escala.	4.1. Modelos de aprendizaje supervisado con Big Data. 4.2. Detección de datos atípicos. 4.3. Librerías, frameworks y herramientas.
5	Aprendizaje no supervisado a gran escala.	5.1. Modelos de aprendizaje no supervisado con Big Data. 5.2. Reducción de dimensiones. 5.3. Librerías, frameworks y herramientas.
6	Minería de texto a gran escala.	6.1 Indexación semántica latente. 6.2 Modelado temático. 6.3 Asignación latente Dirichlet. 6.4 Librerías de procesamiento de lenguaje natural.

7. Actividades de aprendizaje de los temas

1. Introducción a Big Data	
Competencias	Actividades de aprendizaje
Específica(s): Domina los conceptos fundamentales del procesamiento con Big Data, incluyendo la infraestructura tecnológica, herramientas y técnicas básicas para el manejo y análisis de grandes volúmenes de datos. *Genérica(s):* ● Capacidad de análisis y síntesis: Identificar y comprender los elementos clave del procesamiento con Big Data, así como sintetizar información compleja para su aplicación en diferentes contextos. ● Habilidades de investigación: Buscar, recopilar y evaluar información relevante sobre tecnologías y metodologías de procesamiento con Big Data. ● Trabajo en equipo: Colaborar con compañeros en actividades prácticas para resolver problemas y realizar proyectos relacionados con el procesamiento con Big Data.	● Lecturas y estudio dirigido: lectura de material teórico sobre los fundamentos del procesamiento con Big Data, incluyendo conceptos, arquitecturas y tecnologías clave. ● Prácticas en laboratorio: Realización de ejercicios prácticos utilizando herramientas como Hadoop, Spark o sistemas de bases de datos distribuidas para familiarizarse con el manejo de grandes volúmenes de datos. ● Análisis de casos: Estudio y análisis de casos de estudio reales donde se aplicaron técnicas de procesamiento con Big Data para resolver problemas específicos en diferentes industrias. ● Debates y discusiones: Participación en debates y discusiones en clase sobre temas relevantes relacionados con el procesamiento con Big Data, como desafíos éticos, impacto en la sociedad y tendencias futuras. ● Proyecto de investigación: Desarrollo de un proyecto de investigación o aplicación práctica que involucre el uso de herramientas y técnicas de procesamiento con Big Data, desde la recolección y preparación de datos hasta el análisis y la presentación de resultados.
2. Procesamiento a gran escala	
Competencias	Actividades de aprendizaje
Específica(s): Implementa infraestructuras para el procesamiento eficiente de grandes volúmenes de datos. *Genérica(s):* ● Habilidades de trabajo en equipo: Fomentar la colaboración en la resolución de problemas y la realización de proyectos prácticos. ● Habilidades de comunicación: Desarrollar la capacidad de expresar ideas de manera clara y efectiva, tanto de forma oral como escrita.	● Construcción de infraestructuras de datos distribuidas: los estudiantes pueden participar en la configuración y despliegue de sistemas de procesamiento distribuido, como Apache Hadoop o Apache Spark, en entornos de laboratorio simulados o en la nube. ● Implementación de pipelines de datos: los estudiantes pueden trabajar en equipos para diseñar y desarrollar pipelines de datos escalables que abarquen desde la ingestión de datos hasta su procesamiento y almacenamiento en sistemas distribuidos. ● Análisis y optimización de rendimiento: los estudiantes pueden realizar análisis de

<table>
<tr><td>

- Pensamiento crítico: Estimular el análisis reflexivo y la evaluación de diferentes enfoques para abordar problemas de procesamiento de datos a gran escala.
- Capacidad de aprendizaje autónomo: Promover la capacidad de los estudiantes para investigar, adquirir nuevos conocimientos y aplicarlos de manera independiente.

</td><td>

rendimiento y optimización en infraestructuras de procesamiento a gran escala, identificando cuellos de botella y aplicando técnicas para mejorar la eficiencia y la velocidad de procesamiento.
- Resolución de problemas prácticos: los estudiantes pueden enfrentarse a desafíos prácticos relacionados con el procesamiento de datos a gran escala, como la manipulación de conjuntos de datos masivos y la implementación de algoritmos distribuidos para tareas específicas de análisis.
- Presentaciones y discusiones: los estudiantes pueden realizar presentaciones sobre casos de estudio o proyectos relacionados con el procesamiento a gran escala, y participar en discusiones grupales para compartir ideas, experiencias y mejores prácticas.

</td></tr>
<tr><td colspan="2">3. Optimización y modelado de datos</td></tr>
<tr><td>Competencias</td><td>Actividades de aprendizaje</td></tr>
<tr><td>

Específica(s):
Optimiza procesos de modelado de datos, tanto en contextos supervisados como no supervisados, para la generación de insights significativos.

Genérica(s):

- Pensamiento crítico: Fomentar la capacidad de analizar y evaluar diferentes enfoques para la optimización y modelado de datos, así como para la interpretación de resultados.
- Resolución de problemas: Desarrollar habilidades para identificar y abordar eficazmente los desafíos relacionados con la optimización y el modelado de datos.

</td><td>

- Exploración y preprocesamiento de datos: los estudiantes pueden trabajar en la exploración inicial de conjuntos de datos, identificando posibles problemas de calidad de datos, outliers y características relevantes para el modelado.
- Selección y optimización de algoritmos: los estudiantes pueden comparar y evaluar diferentes algoritmos de modelado de datos en función de métricas de desempeño específicas, como precisión, recall o error cuadrático medio, y seleccionar aquellos más adecuados para sus datos y objetivos.
- Validación de modelos: los estudiantes pueden aprender técnicas de validación de modelos, como validación cruzada y división de conjuntos de entrenamiento/prueba, para evaluar la capacidad predictiva y generalización de los modelos desarrollados.

</td></tr>
</table>

• Comunicación efectiva: Promover la capacidad de expresar ideas y resultados de manera clara y coherente, tanto de forma oral como escrita. • Trabajo en equipo: Fomentar la colaboración en proyectos prácticos que involucren la optimización y el modelado de datos, promoviendo la diversidad de ideas y enfoques.	• Taller de optimización de hiper parámetros: los estudiantes pueden participar en talleres prácticos donde aprendan a ajustar los hiper parámetros de los modelos de manera efectiva para mejorar su rendimiento y evitar el sobreajuste. • Desarrollo de informes y presentaciones: los estudiantes pueden elaborar informes y presentaciones que documenten sus procesos de optimización y modelado de datos, incluyendo análisis de resultados, interpretaciones y recomendaciones para la toma de decisiones.
4. Aprendizaje supervisado a gran escala	
Competencias	Actividades de aprendizaje
Específica(s): Aplica técnicas de aprendizaje supervisado a gran escala para la identificación de patrones y la realización de predicciones precisas en conjuntos de datos masivos. *Genéricas:* • Pensamiento crítico: Fomentar la capacidad de evaluar diferentes enfoques de aprendizaje supervisado y comprender sus implicaciones en el análisis de datos a gran escala. • Habilidades de resolución de problemas: Desarrollar la capacidad de identificar y abordar eficazmente desafíos relacionados con el aprendizaje supervisado en grandes conjuntos de datos. • Comunicación efectiva: Promover la capacidad de comunicar resultados de manera clara y coherente, tanto oralmente como por escrito, a diferentes audiencias. • Trabajo en equipo: Fomentar la colaboración en proyectos prácticos que involucren el aprendizaje	• Exploración de algoritmos de aprendizaje supervisado: los estudiantes pueden estudiar y comparar diferentes algoritmos de aprendizaje supervisado, como regresión lineal, regresión logística, árboles de decisión, máquinas de vectores de soporte (SVM) y redes neuronales, para comprender sus principios y aplicaciones en grandes conjuntos de datos. • Implementación y evaluación de modelos: los estudiantes pueden trabajar en equipos para implementar y entrenar modelos de aprendizaje supervisado en conjuntos de datos masivos, utilizando herramientas como Pyspark, y evaluar el rendimiento de los modelos utilizando métricas relevantes, como precisión, recall y F1-score. • Optimización de modelos: los estudiantes pueden aprender técnicas de optimización de modelos, como la selección de características, la optimización de hiper parámetros y la regularización, para mejorar el rendimiento y la generalización de los modelos de aprendizaje supervisado. • Validación cruzada y evaluación de rendimiento: los estudiantes pueden aplicar técnicas de validación cruzada y división de conjuntos de entrenamiento/prueba para evaluar el rendimiento de los modelos de aprendizaje supervisado en

supervisado a gran escala, aprovechando la diversidad de habilidades y experiencias del equipo.	grandes conjuntos de datos y evitar el sobreajuste. ● Análisis e interpretación de resultados: los estudiantes pueden analizar y interpretar los resultados obtenidos de los modelos de aprendizaje supervisado, identificando patrones y tendencias significativas que puedan utilizarse para la toma de decisiones informadas.
5. Aprendizaje no supervisado a gran escala	
Competencias	Actividades de aprendizaje
Específica(s): Aplica técnicas de aprendizaje no supervisado a gran escala para la identificación de patrones y estructuras ocultas en conjuntos de datos masivos. *Genéricas:* ● Pensamiento crítico: Fomentar la capacidad de evaluar diferentes enfoques de aprendizaje no supervisado y comprender sus implicaciones en el análisis de datos a gran escala. ● Habilidades de resolución de problemas: Desarrollar la capacidad de identificar y abordar eficazmente desafíos relacionados con el aprendizaje no supervisado en grandes conjuntos de datos. ● Comunicación efectiva: Promover la capacidad de comunicar resultados de manera clara y coherente, tanto oralmente como por escrito, a diferentes audiencias. ● Trabajo en equipo: Fomentar la colaboración en proyectos prácticos que involucren el aprendizaje no supervisado a gran escala, aprovechando la diversidad de	● Exploración de algoritmos de aprendizaje supervisado: los estudiantes pueden estudiar y comparar diferentes algoritmos de aprendizaje supervisado, como regresión lineal, regresión logística, árboles de decisión, máquinas de vectores de soporte (SVM) y redes neuronales, para comprender sus principios y aplicaciones en grandes conjuntos de datos. ● Implementación y evaluación de modelos: los estudiantes pueden trabajar en equipos para implementar y entrenar modelos de aprendizaje supervisado en conjuntos de datos masivos, utilizando herramientas como Pyspark, y evaluar el rendimiento de los modelos utilizando métricas relevantes, como precisión, recall y F1-score. ● Optimización de modelos: los estudiantes pueden aprender técnicas de optimización de modelos, como la selección de características, la optimización de hiper parámetros y la regularización, para mejorar el rendimiento y la generalización de los modelos de aprendizaje supervisado. ● Validación cruzada y evaluación de rendimiento: los estudiantes pueden aplicar técnicas de validación cruzada y división de conjuntos de entrenamiento/prueba para evaluar el rendimiento de los modelos de aprendizaje supervisado en

habilidades y experiencias del equipo.	grandes conjuntos de datos y evitar el sobreajuste. ● Análisis e interpretación de resultados: los estudiantes pueden analizar y interpretar los resultados obtenidos de los modelos de aprendizaje supervisado, identificando patrones y tendencias significativas que puedan utilizarse para la toma de decisiones informadas.
6. Minería de texto a gran escala	
Competencias	Actividades de aprendizaje
Específica(s): Aplica técnicas de minería de texto a gran escala para extraer información relevante y significativa de documentos y fuentes de datos textuales. *Genéricas:* ● Pensamiento crítico: Fomentar la capacidad de evaluar diferentes enfoques de aprendizaje supervisado y comprender sus implicaciones en el análisis de datos a gran escala. ● Habilidades de resolución de problemas: Desarrollar la capacidad de identificar y abordar eficazmente desafíos relacionados con el aprendizaje supervisado en grandes conjuntos de datos. ● Comunicación efectiva: Promover la capacidad de comunicar resultados de manera clara y coherente, tanto oralmente como por escrito, a diferentes audiencias. ● Trabajo en equipo: Fomentar la colaboración en proyectos prácticos que involucren el aprendizaje supervisado a gran escala, aprovechando la diversidad de	● Preprocesamiento de texto: Los estudiantes pueden realizar tareas de preprocesamiento de texto en grandes conjuntos de datos, como tokenización, eliminación de stopwords, lematización o stemming, para preparar los datos para su análisis. ● Vectorización de texto: Los estudiantes pueden aprender técnicas de vectorización de texto, como la representación TF-IDF (Term Frequency-Inverse Document Frequency) o la incrustación de palabras (word embeddings), para convertir documentos de texto en representaciones numéricas que puedan ser utilizadas por algoritmos de aprendizaje automático. ● Análisis de sentimientos: Los estudiantes pueden aplicar técnicas de análisis de sentimientos para identificar la polaridad y la subjetividad de textos en grandes conjuntos de datos, como comentarios de redes sociales o reseñas de productos. ● Clasificación de texto: Los estudiantes pueden implementar algoritmos de clasificación de texto, como el clasificador Naive Bayes o máquinas de vectores de soporte (SVM), para categorizar automáticamente documentos de texto en diferentes clases o categorías. ● Agrupamiento de texto: Los estudiantes pueden investigar y aplicar algoritmos de agrupamiento de texto, como k-means o Latent Dirichlet Allocation (LDA), para identificar temas o tópicos emergentes en grandes colecciones de documentos de texto.

	habilidades y experiencias del equipo.	

8. Práctica(s)

- Implementación de un sistema de procesamiento distribuido: Los estudiantes podrían trabajar en equipos para configurar y desplegar un sistema de procesamiento distribuido utilizando Apache Hadoop o Apache Spark en un entorno de laboratorio. Esto incluiría la instalación y configuración de los componentes necesarios, como HDFS (Hadoop Distributed File System) on Spark Cluster.
- Análisis de datos en tiempo real: Los estudiantes podrían desarrollar una aplicación de procesamiento de datos en tiempo real utilizando Apache Kafka y Apache Spark Streaming. Esta práctica implicaría la ingestión de datos en tiempo real desde diferentes fuentes, como logs de servidores web o redes sociales, y el análisis en tiempo real para la detección de patrones o tendencias.
- Desarrollo de pipelines de procesamiento de datos: Los estudiantes podrían diseñar y desarrollar pipelines de procesamiento de datos completos que abarquen desde la ingestión de datos hasta su análisis y visualización. Esto incluiría la implementación de transformaciones de datos utilizando herramientas como Apache Spark o Apache Beam, y la integración con sistemas de almacenamiento distribuido como HDFS o Apache HBase.
- Análisis y visualización de datos masivos: Los estudiantes podrían realizar análisis y visualización de conjuntos de datos masivos utilizando herramientas como Apache Spark y bibliotecas de visualización como Matplotlib, Seaborn o Plotly. Esta práctica implicaría la exploración y el análisis de grandes volúmenes de datos para identificar patrones, tendencias o anomalías, y la creación de visualizaciones informativas para comunicar los resultados.
- Desarrollo de modelos de machine learning a gran escala: Los estudiantes podrían desarrollar y entrenar modelos de machine learning a gran escala utilizando bibliotecas como Scikit-learn en conjunto con Apache Spark MLlib. Esto incluiría la preparación de los datos, la selección de características, el entrenamiento de los modelos y la evaluación del rendimiento en conjuntos de datos masivos.
- Optimización de rendimiento de procesamiento de datos: Los estudiantes podrían investigar y aplicar técnicas de optimización de rendimiento en sistemas de procesamiento distribuido, como la paralelización de tareas, la optimización de algoritmos o el ajuste de la configuración del clúster. Esto les permitiría mejorar la eficiencia y la velocidad de procesamiento de grandes volúmenes de datos.

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. Erl, T., Khattak, W., & Buhler, P. (2016). Big data fundamentals: concepts, drivers & techniques. Prentice Hall Press.
2. White, T. (2012). Hadoop: The definitive guide. " O'Reilly Media, Inc.".
3. Acharya, S., & Chellappan, S. (2015). Big Data and Analytics. Wiley
4. Kakarla, R., Krishnan, S., & Alla, S. (2021). Applied Data Science Using PySpark: learn the End-to-End Predictive Model-Building Cycle. Apress.