

1. Datos Generales de la asignatura

Nombre de la asignatura:	Minería de Datos y Big Data
Clave de la asignatura:	TDC-2205
SATCA¹:	2-2-4
Carrera:	Ingeniería en Tecnologías de la Información y Comunicaciones

2. Presentación

Caracterización de la asignatura

Esta asignatura aporta al perfil del Ingeniero los conocimientos teóricos-prácticos en el uso de la Minería de datos y Big data, enfatizando en las diferentes tecnologías y aplicaciones de los tipos de bases de datos especializadas y gran escala, para diseñar herramientas para el procesamiento, almacenamiento y transmisión de datos, para su formación y desempeño profesional.

Se relaciona con todas las asignaturas de Estructuras y Organización de Datos, Fundamentos de Bases de Datos, Taller de Base de Datos, Base de Datos Distribuidas y Programación Web.

La Minería de Datos nace de la necesidad de construir patrones inherentes en la posible relación subyacente en los datos, dicha necesidad tiene su origen en la cada vez más creciente capacidad de almacenamiento de los procesos electrónicos en diferentes ámbitos (comerciales, científicos y/o académicos), lo que conlleva al acumulando grandes volúmenes de información. Sin embargo, no toda esta información resulta realmente útil. El encontrar información útil y convertirla en conocimiento no es una tarea fácil. Tradicionalmente, los analistas de datos cumplían esta labor. En la actualidad, el proceso de extracción de información útil requiere la aplicación de técnicas de análisis de datos automáticas que sean capaces de procesar grandes volúmenes de información.

El desarrollo de las redes sociales, las comunidades virtuales y en general toda la información que los usuarios comparten a través de Web ha creado un nuevo campo de aplicación de las técnicas de Análisis Inteligente de Datos.

Big Data es la nueva moda en ingeniería del software. Cuando se acuñó el término, este se utilizaba para técnicas software que necesitaban o eran capaces de trabajar con cantidades de datos que superaran el límite alcanzable por una única máquina. Hoy en día, el término tiene una acepción más amplia como una disciplina que trata de investigar formas de procesamiento digital para la toma de decisiones mediante el uso de activos de información de gran volumen, que cambian rápidamente y son muy variados en forma.

Intención didáctica

El estudio de esta asignatura se divide en cinco temas:

En el primero el estudiante conoce la Minería de datos de cualquier servicio o aplicación y define el funcionamiento de los elementos en que se encuentra, no sólo sirve para delimitar el campo de acción en que se ubica, sino que también ayuda a conocer la forma segura en el procedimiento de ejecución de información.

La unidad dos se centra en el panorama que comprende el concepto de Big Data, el cual no sólo se refiere a grandes volúmenes de datos, si no que abarca las diferentes fuentes de datos estructurados, no estructurados y semi-estructurados que requieren ser almacenados en especializadas bases de datos.

En la tercera unidad el estudiante conocerá los métodos de agregación e interpretará los resultados que se obtengan de usarlos

La cuarta unidad contempla la evaluación de modelos y su estimación

La quinta unidad se centra en la analítica de Big Data, la cual es la aplicación de técnicas de analítica avanzada para operar sobre grandes conjuntos de datos. La analítica es una colección de diferentes tipos de herramientas incluyendo la analítica predictiva, minería de datos, estadística, inteligencia artificial, entre otras. Se abordará cómo se unen los conceptos y se obtienen la analítica de Big Data.

Es importante que el profesor le permita ubicar al estudiante en el panorama general en lo que será su ocupación profesional y los aportes que se dan a la sociedad.

3. Participantes en el diseño y seguimiento curricular del programa

Lugar y fecha de elaboración o revisión	Participantes	Evento
Instituto Tecnológico Superior de Puruándiro, Michoacán. Diciembre 2021	Integrantes de la Academia de Tecnologías de Información y Comunicaciones	Elaboración del programa de estudio equivalente en la Reunión de academia de Tecnologías de la Información y Comunicaciones.

4. Competencia(s) a desarrollar

Competencia(s) específica(s) de la asignatura

- Identifica la infraestructura de un servidor y la forma en que se podrá cumplir con la petición de uno o más usuarios de un servicio de conexión a distancia.
- Identifica la terminología de los diferentes sistemas operativos de un servidor.
- Guarda y recupera la información en distintos medios de almacenamiento.
- Identifica la terminología de las aplicaciones de las bases de datos.
- Conoce y aplica de manera eficiente el mantenimiento de un servidor

5. Competencias previas

- El estudiante deberá dominar herramientas estadísticas y bases de datos.
- Capacidad para análisis e interpretación
- Análisis matemático
- Capacidad de abstracción
- Manejo algebraico de matrices
- Cálculo de determinantes

- Utiliza técnicas de modelado para la solución de problemas.
- Comprende los conceptos de las bases de datos.
- Construye modelos de bases de datos.
- Domina el lenguaje de SQL.
- Diseña, crea y administra esquemas de bases de datos relacionales bajo estándares.
- Conoce distintas plataformas y SGBD.
- Aplica las actividades que involucra cada una de las etapas del ciclo de vida de un proyecto de software.
- Analiza, diseña, desarrolla e implementa un proyecto de software.
- Crea y manipula bases de datos utilizando distintos Gestores de Bases de Datos considerando elementos de integridad y seguridad para el tratamiento de la información en distintas plataformas

6. Temario

No.	Temas	Subtemas
1	Introducción a la Minería de Datos	1.1. Definición de Minería de Datos 1.2. La minería de datos y el proceso de descubrimiento de conocimiento a partir de datos. 1.3. Relación con otras disciplinas 1.4. Fases del proceso de extracción de Conocimiento a partir de datos 1.5. Aplicaciones
2	Introducción al Big data	2.1. Definición e historia del Big Data 2.2. Aspectos legales y tendencias 2.3. Las 5 V de Big Data (Volumen, Variedad, Veracidad, Velocidad, Valor) 2.4. Gobernanza de los datos (Datos abiertos) 2.5. Aplicaciones para el Big Data
3	Tareas y Técnicas de Minería de Datos	3.1. Tareas en minería de datos. 3.2. Correspondencia entre métodos y tareas. 3.3 Caracterización de las técnicas de minería de datos. 3.4 Técnicas de Minería de Datos

		3.4.1. Métodos estadísticos. 3.4.2. Reglas de asociación y dependencia. 3.4.3. Métodos Bayesianos. 3.4.4. Árboles de Decisión y sistemas de reglas. 3.4.5. Máquinas de vectores soporte. 3.4.6. Extracción de conocimiento 3.4.7. Métodos basados en casos y vecindad
4	Evaluación	4.1. Consideraciones generales. 4.2. Técnicas básicas de evaluación de clasificadores 4.2.1. Medidas de la calidad de un clasificador: La tasa de errores. 4.2.2. La descomposición del error en sesgo y varianza: el concepto de generalización 4.2.3. El sobre entrenamiento 4.3 Implantación de un Programa de Minería de Datos (PMD) en una organización 4.3.1. Cuándo implantar un PMD: Necesidades y objetivos 4.3.2. Fases de un PMD: Estándar CRISP- DM. 4.3.3. Integración de un PMD dentro de una organización.
5	Analítica de Big Data	5.1 Tipos de datos 5.1.1 Datos estructurados 5.1.2 Datos no estructurados 5.1.3 Datos semiestructurados 5.2 Fuentes de Big Data 5.3 Transacciones de datos 5.4 Componentes de un Sistema de Big Data 5.4.1 Captura 5.4.2 Transformación 5.4.3 Almacenamiento 5.4.4 Analítica 5.4.5 Visualización 5.4.6 Plataformas 5.5 Analítica de Big Data 5.5.1 Reporting, query y visualización 5.5.2 Analítica predictiva 5.5.3 Analítica Web 5.5.4 Analítica y Listening Social 5.5.5 Analítica M2M

7. Actividades de aprendizaje de los temas

Introducción a la Minería de Datos	
Competencias	Actividades de aprendizaje
<p>Específica(s): Identificar y definir la minería de datos y el concepto de big data Diferenciar entre base de datos y minería de datos</p> <p>Conocer conceptos relacionados con data minig y big data</p> <p>Genéricas:</p> <ul style="list-style-type: none"> • Capacidad de análisis y síntesis • Capacidad de organizar y planificar • Comunicación oral y escrita en su propia lengua • Capacidad de lectura en una segunda lengua • Habilidades de gestión de información (habilidad para buscar y analizar información proveniente de fuentes diversas) • Solución de problemas • Toma de decisiones 	<ul style="list-style-type: none"> • Realizar esquemas de diferenciación y relación de conceptos. • Definir con sus propias palabras que es data minig y big data • Investigar cual ha sido la importancia histórica de estos temas y como ha sido su evolución • Relacionar estos aspectos con el resto de los contenidos de su carrera. • Prospeccionar sobre que se verá más adelante en las asignaturas y realizar un documento de expectativas del curso.
Introducción al Big Data	
Competencias	Actividades de aprendizaje
<p>Específica(s): Comprende la importancia de Big Data, así como sus características y arquitectura.</p> <p>Genéricas:</p> <ul style="list-style-type: none"> • Capacidad de análisis y síntesis • Capacidad de organizar y planificar • Comunicación oral y escrita en su propia lengua • Capacidad de lectura en una segunda lengua • Habilidades de gestión de información (habilidad para buscar y analizar información proveniente de fuentes diversas) • Solución de problemas • Toma de decisiones 	<ul style="list-style-type: none"> • Comprender la definición de Big Data y conocer su importancia. • Conocer la arquitectura básica de Big Data. • Conocer las cuestiones legales para el uso e implementación de Big Data • Conocer los componentes y funciones específicas de Big Data. • Investigar plataformas comerciales de Big Data y sus características. • Proporcionar la capacidad de crear rápidamente paneles de instrumentos y visualizaciones personalizadas.

Tareas y técnicas de minería de datos

Tareas y técnicas de minería de datos	
Competencias	Actividades de aprendizaje
<p>Específica(s): El estudiante aprenderá como se puede evaluar si las técnicas utilizadas en el proceso de minería de datos arrojan resultados significativos.</p> <p>Genéricas:</p> <ul style="list-style-type: none"> • Capacidad de análisis y síntesis • Capacidad de organizar y planificar • Comunicación oral y escrita en su propia lengua • Capacidad de lectura en una segunda lengua • Habilidades de gestión de información (habilidad para buscar y analizar información proveniente de fuentes diversas) • Solución de problemas • Toma de decisiones 	<ul style="list-style-type: none"> • Predicción sobre Series de Tiempo. • Algoritmos Genéticos en Data Mining. Clustering. Sampling. • Active Data Mining. • Web Mining. • Realizar prácticas con alguna herramienta de minería de datos tal como WEKA.
Evaluación	
Competencias	Actividades de aprendizaje
<p>Específica(s): Reflexión y análisis sobre los datos que genera una persona común para la evaluación de datos.</p> <p>Genéricas:</p> <ul style="list-style-type: none"> • Capacidad de análisis y síntesis • Capacidad de organizar y planificar • Comunicación oral y escrita en su propia lengua • Capacidad de lectura en una segunda lengua • Habilidades de gestión de información (habilidad para buscar y analizar información proveniente de fuentes diversas) • Solución de problemas • Toma de decisiones. 	<ul style="list-style-type: none"> • Realizará ejemplos de todo lo que genera datos en nuestra vida y como se clasifican, así como su alcance. • Mediante el uso de estadísticas promedio identificar el nivel de datos que representan a las personas y sus interacciones sociales. • Investigando en diversas fuentes conocer a los personajes, empresas e instituciones que llevan la frontera de conocimiento de big data, así como identificar áreas de investigación posibles en estos temas. • Realizando un trabajo de investigación y tipo ensayo el estudiante imaginará cual será el futuro posible de la minería de datos y el big data, así como relacionará con otros campos de la ciencia y el conocimiento humano.
Análítica de Big Data	
Competencias	Actividades de aprendizaje

<p>Específica(s): Comprende el uso de analítica de Big Data.</p> <p>Genéricas:</p> <ul style="list-style-type: none"> • Capacidad de análisis y síntesis • Capacidad de organizar y planificar • Comunicación oral y escrita en su propia lengua • Capacidad de lectura en una segunda lengua • Habilidades de gestión de información (habilidad para buscar y analizar información proveniente de fuentes diversas) • Solución de problemas • Toma de decisiones 	<ul style="list-style-type: none"> • Conocer la manera de consultar y representar los datos. • Identificar los tipos de datos estructurados, no estructurados y semi-estructurados. • Conocer los componentes y funciones específicas de Big Data. • Conocer e identificar los diferentes tipos de analítica de Big Data.
--	---

8. Práctica(s)

- Manipulación de sistemas operativos de servidores propietario y libre.
- Instalación de software de aplicación de virtualización propietario y libre.
- Instalar y examinar (al menos 3) sistemas operativos Windows Server y Linux.
- Instalar y examinar (al menos 3) buscadores.
- Creación de bases de datos.
- Realización de un coloquio sobre las diversas aplicaciones de los servidores en los diferentes sectores sociales.
- Visitas a diferentes empresas que utilicen los diferentes tipos de servidores.
- Realizar un resumen de visitas a los diferentes sectores que utilicen servidores con las diversas aplicaciones de estos.

9. Proyecto de asignatura

El objetivo del proyecto que plantee 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

Para evaluar las actividades de aprendizaje se recomienda solicitar: cuadros sinópticos, mapas conceptuales, líneas de tiempo, exposiciones en clase, reportes de investigación, portafolio de evidencias, tablas comparativas, entre otros.

Para verificar el nivel del logro de las competencias del estudiante se recomienda utilizar: listas de cotejo, listas de verificación, matrices de valoración, rúbricas, entre otros, mediante la heteroevaluación, coevaluación y autoevaluación.

11. Fuentes de información

1. Margaret H. Dunham -Data Mining: Introductory and Advanced Topics Prentice HallJ.
2. Han and M. Kamber - Data Mining: Concepts and Techniques Morgan Kaufmann Publishers
3. D. Hand, H. Mannila and P. Smyth - Principles of Data Mining MIT Press
4. Ian H. Witten, Eibe Frank - Data Mining: Practical Machine Learning Tools and Techniques with Java Implementations Morgan Kaufmann Publishers
5. Mehmed Kantardzic - Data Mining: Concepts, Models, Methods, and Algorithms Wiley-IEEE Press
6. Tan, Steinbach, Kumar Introduction to Data Mining - Addison-Wesley
7. Usama M. Fayyad, Gregory Piatetsky-Shapiro, Padhraic Smyth Advances in Knowledge Discovery and Data Mining - AAAI/MIT Press
8. Ian H. Witten, Eibe Frank - Data Mining: Practical Machine Learning Tools and Techniques - Morgan Kaufmann, 2nd edition
9. Dorian Pyle - Data Preparation for Data Mining - Morgan Kaufmann
10. Daniel T. Larose - Discovering Knowledge in Data: An Introduction to Data Mining Wiley-Interscience