

## 1. Datos Generales de la asignatura

<b>Nombre de la asignatura:</b>	Minería de Datos y Big Data
<b>Clave de la asignatura:</b>	TEC-1305
<b>SATCA<sup>1</sup>:</b>	2-2-4
<b>Carrera:</b>	Ingeniería en Tecnologías de la Información y Comunicaciones.

## 2. Presentación

### Caracterización de la asignatura

. 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

La asignatura contempla cinco unidades asociadas e interrelacionadas para comprender el proceso de generación de información a partir de grandes colecciones de datos.

En la primera unidad el estudiante comprenderá los aspectos generales de la asignatura y como se recolectan los datos.

La unidad dos el estudiante analizará el cómo se preparan y clasifican los datos para la obtención de información útil.

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 refiere a la implementación e impacto de la minería de datos

<sup>1</sup> Sistema de Asignación y Transferencia de Créditos Académicos

### 3. Participantes en el diseño y seguimiento curricular del programa

Lugar y fecha de elaboración o revisión	Participantes	Observaciones
Instituto Tecnológico Superior de Puruándiro, Michoacán.  Diciembre 2013	Integrantes de la Academia de Tecnologías de Información y Telecomunicaciones	

### 4. Competencia(s) a desarrollar

<b>Competencia(s) específica(s) de la asignatura</b>
Diseñar y aplicar modelos de almacenamiento de datos masivos y en tiempo real. Conocer el proceso de generar nuevo conocimiento a partir de datos y manejar herramientas para el modelado de minería de datos. Ser promotores de una cultura de almacenamiento y registro de datos en diferentes industrias y ámbitos productivos.

### 5. Competencias previas

<ul style="list-style-type: none"> <li>- El estudiante deberá dominar herramientas estadísticas y bases de datos.</li> <li>- Capacidad para análisis e interpretación</li> <li>- Análisis matemático</li> <li>- Capacidad de abstracción</li> <li>- Manejo algebraico de matrices,</li> <li>- cálculo de determinantes</li> </ul>
---

### 6. Temario

No.	Temas	Subtemas
1	<b>Introducción a la minería de datos y Big Data</b>	1.1.El concepto de Minería de Datos 1.2.La minería de datos y el proceso de descubrimiento de conocimiento a partir de datos

		<p>1.3.Relación con otras disciplinas  1.4.Aplicaciones  1.5.Fases del proceso de extracción de conocimiento a partir de datos</p>
<b>2</b>	<b>PREPARACIÓN DE DATOS</b>	<p>2.1.Consideraciones previas generales. Los almacenes de datos.  2.2.Técnicas sencillas de preprocesado  2.2.1.Compleción (datos faltantes)  2.2.2.Limpieza de errores  2.2.3.Transformación de atributos  2.2.4.Escalado  2.2.5.Discretización  2.2.6.Numerización  2.3.Técnicas de reducción de la dimensionalidad I: Análisis de Componentes Principales.  2.4.Técnicas de reducción de la dimensionalidad II: Métodos de Filtrado y Envoltura</p>
<b>3</b>	<b>TAREAS Y TÉCNICAS DE MINERÍA DE DATOS</b>	<p>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.Redes Neuronales Artificiales.  3.4.6.Máquinas de vectores soporte.  3.4.7.Extracción de conocimiento con algoritmos evolutivos y reglas difusas.  3.4.8.Métodos basados en casos y vecindad.</p>
<b>4</b>	<b>EVALUACIÓN</b>	<p>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 sobreentrenamiento  4.2.4.Repetibilidad estadística: la validación cruzada.  4.3.Aspectos específicos de la evaluación de los diferentes clasificadores estudiados  4.4.Técnicas estadísticas de comparación de clasificadores</p>

		<p>4.5. Medidas de calidad de agrupamiento</p> <p>4.6. Interpretación, difusión y uso de modelos</p>
5	<b>IMPLANTACIÓN E IMPACTO BIG DATA</b>	<p>5.1. Implantación de un Programa de Minería de Datos (PMD) en una organización</p> <p>5.1.1. Cuándo implantar un PMD: Necesidades y objetivos</p> <p>5.1.2. Fases de un PMD: Estándar CRISP-DM</p> <p>5.1.3. Integración de un PMD dentro de una organización</p> <p>5.1.4. Recursos necesarios</p> <p>5.2. Repercusiones y retos de la minería de datos</p> <p>5.2.1. Impacto social</p> <p>5.2.2. Cuestiones éticas y legales</p> <p>5.2.3. Problemas y soluciones: Tendencias futuras</p>

### 7. Actividades de aprendizaje de los temas

Nombre de tema	
<b>Introducción a la Minería de datos y Big Data</b>	
Competencias	Actividades de aprendizaje
<p>Específica(s):</p> <p>Identificar y definir la minería de datos y el concepto de big data</p> <p>Diferenciar entre base de datos y minería de datos</p> <p>Conocer conceptos relacionados con data mining y big data</p>	<p>Realizar esquemas de diferenciación y relación de conceptos.</p> <p>Definir con sus propias palabras que es data mining y big data</p> <p>Investigar cual ha sido la importancia histórica de estos temas y como ha sido su evolución</p> <p>Relacionar estos aspectos con el resto de los contenidos de su carrera.</p> <p>Prospectar sobre que se verá más adelante en las asignaturas y realizar un documento de expectativas del curso</p>

Nombre de tema <b>Preparación de Datos</b>	
Competencias	Actividades de aprendizaje
<p>Específica(s):</p> <p>Las habilidades de “tubería de datos”</p> <p>Profundo conocimiento sobre cómo extraer los datos o usar análisis predictivos</p> <p>Matemáticas, modelado y análisis estadístico</p>	<p>Prácticas de filtrado usando alguna herramienta de minería de datos.</p> <p>Prácticas de aprendizaje usando alguna herramienta de minería de datos.</p> <p>Investigar en diversas fuentes de información, las transformaciones que se pueden realizar a los datos de entrada</p> <p>Para facilitar el proceso de minería de datos, hacer un reporte escrito y discutirlo en clase.</p> <p>Realizar ejercicios de agrupamiento (clustering) para dejar más claro su funcionamiento.</p> <p>Realizar ejercicios de construcción de árboles de decisión para dejar más claro su funcionamiento.</p>
Nombre de tema <b>Tareas y técnicas de minería de datos</b>	
Competencias	Actividades de aprendizaje
<p>Específica(s):</p> <p>El estudiante aprenderá como se puede evaluar si las técnicas utilizadas en el proceso de minería de datos arrojan resultados significativos.</p>	<p>Predicción sobre Series de Tiempo.</p> <p>Algoritmos Genéticos en Data Mining.</p> <p>Clustering. Sampling.</p> <p>Active Data Mining.</p> <p>Web Mining.</p> <p>Realizar prácticas con alguna herramienta de minería de datos tal como WEKA.</p>

Nombre de tema <b>Evaluación</b>	
Competencias	Actividades de aprendizaje
<p>Específica(s):</p> <p>Reflexión y análisis sobre los datos que genera una persona común para la evaluación de datos.</p> <p>Conocimiento de la frontera de conocimiento de big data.</p> <p>Prospección y futuro del Big data.</p>	<p>Realizará ejemplos de todo lo que genera datos en nuestra vida y como se clasifican así como su alcance.</p> <p>Mediante el uso de estadísticas promedio identificar el nivel de datos que representan a las personas y sus interacciones sociales.</p> <p>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.</p> <p>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.</p>

Nombre de tema <b>Implantación e Impacto Big Data</b>	
Competencias	Actividades de aprendizaje
<p>Específica(s):</p> <p>Aprenderán habilidades de trabajo para el desarrollo de soluciones de grandes volúmenes de datos.</p> <p>Implementar soluciones para grandes volúmenes de datos.</p>	<p>Indexación universal en tiempo real de cualesquiera datos de máquina</p> <p>Búsqueda libre y análisis de datos históricos y en tiempo real</p> <p>Monitoree sus datos y proporcione alertas en tiempo real</p> <p>Proporcione la capacidad de crear rápidamente paneles de instrumentos y visualizaciones personalizadas</p>

--	--

## 8. Práctica(s)

Recolección de datos usando diferentes fuentes y servicios de redes sociales, desde tweeter a Facebook.

Interpretación y deducción de comportamiento a través de la información recolectada sobre tendencias de y usos.

Generación de algoritmos predictivos de datos.

Investigaciones sobre el uso de los datos y su control, así como leyes que protejas

## 9. Proyecto de asignatura

El estudiante seleccionará un tema de interés para sí mismo o para la comunidad en la que convive, siempre y cuando este tema cuente con datos públicos/privados históricos o de gran volumen o a los cuales tiene acceso para poder ejecutar sobre ellos un análisis y estudio de minería de datos. Por ejemplo datos públicos de estadísticas nacionales de vivienda, salud, agricultura y de economía pueden ser analizados desde una perspectiva de minería de datos para generar nuevo conocimiento. O bien temas privados en donde existen datos históricos y de gran volumen como el béisbol, bolsa de valores o redes sociales pueden ser sujetos a un análisis de minería de datos.

Al final se unificarán cada una de las prácticas de unidad y se modelará en concordancia a obtener un resultado incremental con el proyecto, o bien, que ayude a que pueda realizarse la misma práctica con los datos obtenidos.

El estudiante obtendrá y modelará los datos para su análisis en minería de datos con fines de obtener conocimiento nuevo o bien confirmar y tener evidencia del ya existente.

## 10. Evaluación por competencias

- Solución de casos prácticos solicitados durante las actividades, así como sus conclusiones de forma escrita.
- Reportes de investigación de campo.
- Reportes de prácticas.
- Ejercicios realizados.
- Tareas
- Exposición.
- Participación en clase.
- Proyecto.
- Exámenes teóricos y/o prácticos.

## 11. Fuentes de información

1. Margaret H. Dunham -Data Mining: Introductory and Advanced Topics Prentice Hall
2. J. 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