



"2020, Año de Leona Vicario, Benemérita Madre de la Patria"

1. DATOS GENERALES DE LA ASIGNATURA

| | |
|---------------------------------|---|
| Nombre de la asignatura: | Minería de Datos |
| Clave de la asignatura: | CDF-2002 |
| SATCA¹: | 3-2-5 |
| Carrera: | Ingeniería en Sistemas Computacionales |

2. PRESENTACIÓN

Caracterización de la asignatura

La materia de Minería de Datos (DM) proporciona al estudiante una introducción a los métodos de extracción de información útil para la toma de decisiones a partir de un conjunto de datos estructurados o sin estructura y que son procedentes de distintas fuentes de información no necesariamente homogéneas. DM representa una de las tres principales técnicas para la Ciencia de Datos que ejecuta el proceso de convertir los datos en información y a esta en conocimiento.

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 de 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. Estos métodos fueron originados en áreas de computación, Probabilidad, Estadística e Inteligencia Artificial.

La Minería de Datos se define como el proceso de aplicar metodologías basadas en computadores, conceptos estadísticos y administración para extraer conocimiento útil de grandes volúmenes de información. Actualmente y de forma muy general, los métodos permiten aproximar funciones lineales y no lineales; asimismo clasificar, segmentar, extraer patrones y tendencias a partir de un grupo de datos. Los métodos de la Minería de Datos han sido aprovechados por las instituciones proveedoras de productos y servicios para analizar, organizar y comprender la información generada en operaciones de diversa índole. De tal forma que se ha permitido apoyar la toma de decisiones en organizaciones, por ejemplo, vinculadas a la extracción petrolera, los servicios financieros y de crédito, las transacciones electrónicas u observaciones científicas.

¹ Sistema de Asignación y Transferencia de Créditos Académicos



"2020, Año de Leona Vicario, Benemérita Madre de la Patria"

Intención didáctica

El propósito de este curso es proporcionar una visión panorámica de la Minería de Datos; qué tipo de problemas puede resolver, en qué consisten sus técnicas, qué herramientas existen y cómo se desarrolla una aplicación de Minería de Datos. Para conseguir el objetivo, la realización de trabajos, ejercicios, prácticas y programas por parte de los estudiantes constituye un elemento importante del aprendizaje.

En la primera unidad, se presenta el panorama general de la Minería de Datos, se definen los conceptos básicos, su relación con el proceso de descubrimiento de información, sus aplicaciones, las implicaciones éticas, así como las herramientas de software desarrolladas para la aplicación de esta disciplina.

En la segunda unidad, se describen las etapas previas al procesamiento de información derivadas de la variabilidad de los distintos formatos que pudieran presentar las fuentes de datos. Se explican conceptos tales como: la limpieza de los datos, la integración y transformación de los datos, discretización y jerarquías conceptuales, y por último la reducción de datos.

En la tercera unidad, se estudian los métodos básicos para el análisis de datos con el fin de extraer patrones de interés. Esta unidad organiza los métodos en dos categorías principales; los métodos basados en técnicas en Probabilidad y Estadística; y la segunda categoría engloba los métodos basados en técnicas simbólicas aplicadas a la Minería de Datos.

La cuarta unidad, se enfoca a la evaluación y difusión de los patrones encontrados, donde se examinan las técnicas de evaluación, el grado de interés de los patrones, su aplicación, la transformación de estos patrones en reglas comprensibles y por último la interpretación, difusión y explotación de los modelos encontrados.

Finalmente, en la quinta unidad, se exponen situaciones relevantes sobre casos de estudio en los que la Minería de Datos ha funcionado exitosamente. El objetivo, es relacionar o implementar los conceptos vistos en unidades anteriores a un problema de la industria o del ámbito científico.

Con esta materia, el Ingeniero en Sistemas Computacionales tendrá las herramientas necesarias para poder interactuar de manera congruente con profesionales en otros ámbitos de la ciencia para solucionar problemas con bases científico-metodológicas congruentes, afrontando los retos actuales del desarrollo tecnológico y económico.

Las competencias generales que pueden estimularse son, entre otras:

- Adquirir la capacidad de discernir los aspectos relevantes de investigaciones en problemas reales.



"2020, Año de Leona Vicario, Benemérita Madre de la Patria"

- Adquirir la capacidad de análisis y síntesis de problemas de procesamiento de información.
- Integración de grupos de trabajo, en ocasiones multidisciplinarios.
- Aplicar conocimientos adquiridos en la solución de problemas a planteamientos específicos.
- Uso de herramientas informáticas para el desarrollo e implementación de soluciones.

El profesor que imparta esta asignatura debe:

- Propiciar actividades de búsqueda, selección y análisis de información en distintas fuentes.
- Propiciar el uso de las nuevas tecnologías en el desarrollo de los temas de la asignatura.
- Propiciar actividades de planeación y organización de distinta índole en el desarrollo de la asignatura.
- Fomentar actividades grupales que propicien la comunicación, el intercambio argumentado de ideas, la reflexión, la integración, y la colaboración de y entre los estudiantes.
- Propiciar, en el estudiante, el desarrollo de actividades intelectuales de inducción-deducción y análisis-síntesis, las cuales lo encaminan hacia la investigación, la aplicación de conocimientos y la solución de problemas.
- Llevar a cabo actividades prácticas que promuevan el desarrollo de habilidades para la experimentación, tales como: observación, identificación manejo y control de variables y datos relevantes, planteamiento de hipótesis, de trabajo en equipo.
- Desarrollar actividades de aprendizaje que propicien la aplicación de los conceptos, modelos y metodologías que se van aprendiendo en el desarrollo de la asignatura.
- Propiciar el uso adecuado de conceptos, y de terminología científico-tecnológica
- Proponer problemas que permitan al estudiante la integración de contenidos de la asignatura y entre distintas asignaturas, para su análisis y solución.
- Relacionar los contenidos de la asignatura con el cuidado del medio ambiente; así como con las prácticas de una carrera técnica con enfoque sustentable.
- Observar y analizar fenómenos y problemáticas propias del campo ocupacional.



"2020, Año de Leona Vicario, Benemérita Madre de la Patria"

3. PARTICIPANTES EN EL DISEÑO Y SEGUIMIENTO CURRICULAR DEL PROGRAMA

| Lugar y fecha de elaboración o revisión | Participantes | Observaciones |
|--|--|---|
| Instituto Tecnológico de Acapulco. Mayo del 2020 | Academia de Sistemas y Computación Dra. Miriam Martínez Arroyo. Dr. José Antonio Montero Valverde. M.C. José Francisco Gazga Portillo. Dra. Carolina Astudillo Hernández | Definición de los programas de estudio de especialidad para la carrera de Ingeniería en Sistemas Computacionales. |

4. COMPETENCIA(S) A DESARROLLAR

| Competencia(s) específica(s) de la asignatura |
|---|
| El estudiante comprende los conceptos básicos de la Minería de Datos, el proceso que se sigue para la preparación de los datos, las técnicas de análisis exploratorio y finalmente la evaluación, difusión y aplicabilidad de los resultados obtenidos destinados a resolver problemas del mundo real, para conseguir esto, en el transcurso de la materia se elaboran resúmenes, se aplican exámenes y se desarrollan prácticas de laboratorio, lo anterior facilitará el reconocimiento de las fases del proceso de descubrimiento de información en bases de datos así como de los algoritmos y técnicas aplicables a la Minería de Datos. |

5. COMPETENCIAS PREVIAS

| |
|---|
| <ul style="list-style-type: none"> • Diseña, Interpreta e implementa algoritmos computacionales y notaciones matemáticas a través del uso de herramientas computacionales adecuadas tales como: software para modelado y lenguajes de programación. • Diseña, Interpreta e implementa sistemas de Bases de Datos mediante el uso de software de gestión de bases de datos. • Comprende los conceptos de Probabilidad, Cálculos Estadísticos, Redes de Bayes, Correlación, Regresión Lineal, Teoría de Conjuntos, Redes Neuronales y Lógica de Predicados y Difusa mediante su aplicación a la resolución de problemas prácticos. |
|---|



"2020, Año de Leona Vicario, Benemérita Madre de la Patria"

6. TEMARIO

| No. | Temas | Subtemas |
|-----|--|--|
| 1 | Visión general de la Minería de Datos | 1.1. La Minería de Datos y el KDD 1.2. Comparativa entre el Aprendizaje Automático y la Minería de Datos 1.3. Software para Minería de Dato |
| 2 | Preprocesamiento de datos | 2.1 Limpieza de datos 2.2 Integración y Transformación 2.3 Discretización de datos y Jerarquías Conceptuales 2.4 Reducción de datos |
| 3 | Métodos básicos para el análisis de datos | 3.1. Procedimientos estadísticos aplicados para la minería de datos 3.1.1. Técnicas para agrupamiento 3.1.2. Técnicas para clasificación 3.2. Procedimientos simbólicos aplicados para la minería de datos 3.2.1. Computación evolutiva 3.2.2. Razonamiento basado en casos |
| 4 | Evaluación y difusión de patrones encontrados | 4.1. Técnicas de evaluación 4.2. Comprensibilidad, Interés y Aplicabilidad 4.3. Extracción de reglas comprensibles 4.4. Interpretación, difusión y uso de Modelos |
| 5 | Aplicación de la Minería de Datos (Caso Práctico) | 5.1. Planteamiento 5.2. Preprocesamiento de datos 5.3. Definición, parametrización e implementación del algoritmo de Minería 5.4. Análisis, visualización y validación de Resultados 5.5. Conclusiones |



"2020, Año de Leona Vicario, Benemérita Madre de la Patria"

7. ACTIVIDADES DE APRENDIZAJE DE LOS TEMAS

| Tema 1.- Visión general de la Minería de Datos | |
|--|--|
| Competencias | Actividades de aprendizaje |
| <p>Específica(s):</p> <ul style="list-style-type: none"> • Conoce los conceptos y características básicas de la Minería de Datos, además de reconocer y relacionar temas de otras áreas de conocimiento como base de la Minería de Datos con el fin de ser tomadas en cuenta en su aplicación y desarrollo de soluciones construyendo cada una de las fases del KDD para la toma de decisiones. <p>Genéricas:</p> <ul style="list-style-type: none"> • Aplica actividades de búsqueda, selección y análisis de información en distintas fuentes. • Realiza actividades grupales que propicien la comunicación, el intercambio argumentado de ideas, la reflexión, la integración y la colaboración de y entre los estudiantes. • Resuelve problemas que la integración de contenidos de la asignatura y entre distintas asignaturas, para su análisis y solución. • Usa la terminología técnica adecuada al programa. | <ul style="list-style-type: none"> • Definir y comprender el campo de la Minería de Datos. • Conocer el proceso de KDD. • Aplicar conocimientos de Base de Datos para comprender el alcance y áreas de oportunidad de la Minería de Datos. • Investigar software de propósito específico para el procesamiento y análisis de información aplicando algoritmos de Minería de Datos. |
| Tema 2.- Preprocesamiento de datos | |
| Competencias | Actividades de aprendizaje |
| <p>Específica(s):</p> <ul style="list-style-type: none"> • Conoce y aplica las actividades a realizar en la limpieza, integración, transformación, discretización y reducción de datos, con el fin de construir una base de datos homogénea a partir de la cual monitorizar el comportamiento de los datos. <p>Genéricas:</p> <ul style="list-style-type: none"> • Aprende nuevas herramientas y técnicas para implementar modelos | <ul style="list-style-type: none"> • Estudiar las etapas del preprocesamiento. • Aplicar una o más técnicas de cada una de las etapas del preprocesamiento a fuentes de datos de diferente naturaleza. • Construir bases de datos homogéneas a partir de diferentes fuentes de información. • Utilizar software de propósito específico para el procesamiento y análisis de |



"2020, Año de Leona Vicario, Benemérita Madre de la Patria"

| | |
|--|--|
| <p>estadísticos, probabilistas y de toma de decisiones.</p> <ul style="list-style-type: none"> • Aplica actividades de búsqueda, selección y análisis de información en distintas fuentes. • Realiza actividades grupales que propicien la comunicación, el intercambio argumentado de ideas, la reflexión, la integración y la colaboración de y entre los estudiantes. • Desarrolla actividades de aprendizaje que propicien la aplicación de los conceptos, modelos y metodologías que se van aprendiendo en el desarrollo de la asignatura. • Utiliza las nuevas tecnologías en el desarrollo de la asignatura. • Usa la terminología técnica adecuada al programa. • Realiza la exposición en plenaria de un proyecto final con apoyo de material didáctico (cañón, proyector de acetatos, rotafolios+, pizarrón, entre otros). | <p>información aplicando algoritmos de Minería de Datos.</p> |
|--|--|

Tema 3.- Métodos básicos para el análisis de datos

| Competencias | Actividades de aprendizaje |
|---|--|
| <p>Específica(s):</p> <ul style="list-style-type: none"> • Conoce, diferencia y aplica los métodos básicos para el análisis de datos que se aplican en la Minería de Datos, con el fin de obtener patrones de interés. • Genera los modelos y patrones utilizando las herramientas de Minería de Datos disponibles. <p>Genéricas:</p> <ul style="list-style-type: none"> • Aprende nuevas herramientas y técnicas para implementar modelos estadísticos, probabilistas y de toma de decisiones. • Aplica actividades de búsqueda, selección y análisis de información en distintas fuentes. | <p>Estudiar los procedimientos estadísticos y simbólicos para el análisis de datos.</p> <p>Identificar las condiciones bajo las cuales aplicar técnicas estadísticas o simbólicas.</p> <p>Aplicar las diversas técnicas de análisis a un caso de estudio.</p> <p>Utilizar software de propósito específico para el procesamiento y análisis de información aplicando algoritmos de Minería de Datos.</p> |



"2020, Año de Leona Vicario, Benemérita Madre de la Patria"

| | |
|---|--|
| <ul style="list-style-type: none"> • Realiza actividades grupales que propicien la comunicación, el intercambio argumentado de ideas, la reflexión, la integración y la colaboración de y entre los estudiantes. • Desarrolla actividades de aprendizaje que propicien la aplicación de los conceptos, modelos y metodologías que se van aprendiendo en el desarrollo de la asignatura. • Propiciar el uso de las nuevas tecnologías en el desarrollo de la asignatura. • Usa la terminología técnica adecuada al programa. • Realiza la exposición en plenaria de un proyecto final con apoyo de material didáctico (cañón, proyector de acetatos, rotafolios, pizarrón, entre otros). | |
| <p>Tema 4.- Evaluación y difusión de patrones encontrados</p> | |
| <p>Competencias</p> | <p>Actividades de aprendizaje</p> |
| <p>Específica(s):</p> <ul style="list-style-type: none"> • Evalúa la calidad de los patrones resultantes de la aplicación de algún algoritmo de Minería con el fin de determinar la comprensibilidad, interés y aplicabilidad del modelo, así como también su interpretación y difusión para la toma de decisiones. <p>Genéricas:</p> <ul style="list-style-type: none"> • Aplica actividades de búsqueda, selección y análisis de información en distintas fuentes. • Realiza actividades grupales que propicien la comunicación, el intercambio argumentado de ideas, la reflexión, la integración y la colaboración de y entre los estudiantes. • Desarrolla actividades de aprendizaje que propicien la aplicación de los conceptos, modelos y metodologías | <ul style="list-style-type: none"> • Estudiar las distintas técnicas de evaluación de patrones. • Aplicar las diversas técnicas de evaluación de patrones a un caso de estudio. • Utilizar software de propósito específico para el procesamiento y análisis de información aplicando algoritmos de Minería de Datos. |



"2020, Año de Leona Vicario, Benemérita Madre de la Patria"

| | |
|--|--|
| <p>que se van aprendiendo en el desarrollo de la asignatura.</p> <ul style="list-style-type: none"> • Utiliza las nuevas tecnologías en el desarrollo de la asignatura. • Usa la terminología técnica adecuada al programa. | |
| <p>Tema 5.- Aplicación de la Minería de Datos (Caso Práctico)</p> | |
| <p>Competencias</p> | <p>Actividades de aprendizaje</p> |
| <p>Específica(s):</p> <ul style="list-style-type: none"> • Descubre un patrón o modelo que se encuentra oculto en un cmulo significativo de datos que provienen de distinta fuente, mediante la Minería de Datos, con el fin de soportar y favorecer la toma de decisiones. <p>Genéricas:</p> <ul style="list-style-type: none"> • Desarrolla la capacidad para coordinar y trabajar en equipo; orientar el trabajo y potenciar la autonomía, el trabajo cooperativo y la toma de decisiones. • Presentan un proyecto de asignatura que incluya los aspectos relevantes de la investigación de campo. El proyecto incluye una presentación escrita y una oral. Todos los integrantes de cada grupo de trabajo deben participar para incentivar y promover el desarrollo de estas capacidades. • Realiza actividades grupales que propicien la comunicación, el intercambio argumentado de ideas, la reflexión, la integración y la colaboración de y entre los estudiantes. • Utiliza las nuevas tecnologías en el desarrollo de la asignatura. • Realiza la exposición en plenaria de un proyecto final con apoyo de material didáctico (cañón, proyector de acetatos, rota folios, pizarrón, entre otros). | <ul style="list-style-type: none"> • Aplicar los conceptos estudiados en las unidades previas a un caso de estudio. • Aplicar técnicas de preprocesamiento de datos, análisis de datos, evaluación de patrones para tomar decisiones sobre un problema que se presente en algún ámbito de estudio (académico, científico, laboral, social, industrial, político, electoral, criminal, etc.) • Utilizar software de propósito específico para el procesamiento y análisis de información aplicando algoritmos de Minería de Datos. |



"2020, Año de Leona Vicario, Benemérita Madre de la Patria"

8. PRÁCTICA(S)

- Investigar aplicaciones de la Minería de Datos y el software especializado para tal fin. Además, aplicar los conceptos de Limpieza, Integración y Transformación de Datos.
- Aplicar métodos descriptivos para el análisis de información tales como: Agrupamiento, Reglas de Asociación y/o Correlación.
- Aplicar métodos predictivos para el análisis de información tales como: Clasificación y/o Regresión.
- Determinar la calidad y/o grado de interés, aplicabilidad y comprensión de patrones encontrados en bases de datos.
- Aplicar el proceso del KDD para tomar decisiones sobre un problema que se presente en algún ámbito de estudio (académico, científico, laboral, social, industrial, político, electoral, criminal, etc.)

9. PROYECTO DE ASIGNATURA.

Descubrir un modelo o patrón descriptivo o predictivo inherente y oculto en un conjunto basto de datos procedentes de distintas fuentes de información, con el fin de difundir y aplicar dicho patrón para favorecer la toma de decisiones. El proyecto deberá cumplir con las siguientes características:

- **Fundamentación:** Marco referencial (teórico, conceptual, contextual, legal) en el cual se fundamenta el proyecto de acuerdo con un diagnóstico realizado, con el fin de 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, esto implica planificar un proceso: de intervención empresarial, social o comunitaria, 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, 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:** Aplicar 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.



"2020, Año de Leona Vicario, Benemérita Madre de la Patria"

10. EVALUACIÓN POR COMPETENCIAS

Son las técnicas, instrumentos y herramientas sugeridas para constatar los desempeños académicos de las actividades de aprendizaje.

- Aplicar 1 evaluaciones diagnóstica para verificar las competencias previas.
- Aplicar 2 evaluaciones formativas y al final una evaluación sumativa.
- Realizar investigaciones documentales referentes a la asignatura usando los diferentes medios bibliográficos o electrónicos, para desarrollar posteriormente: cuadros comparativos, mapas conceptuales, cuadros sinópticos, resúmenes y ensayos.
- Representar, comparar, reflexionar sobre teorías o conceptos.
- Realizar prácticas y ejercicios en los diferentes tópicos de la asignatura.
- Evaluar el desempeño del estudiante en el grupo utilizando instrumentos de autoevaluaciones y coevaluaciones (por ejemplo: rúbricas o listas de cotejo).
- Delimitar las especificaciones de los proyectos.
- Desarrollar proyectos usando herramientas gráficas, donde se aplique el manejo de métodos probabilísticos.
- Evaluar el diseño e implementación del proyecto final.
- Ponderar tareas.
- Cumplimiento de los objetivos y desempeño en las prácticas



"2020, Año de Leona Vicario, Benemérita Madre de la Patria"

11. FUENTES DE INFORMACIÓN

1. Hernández Orallo, J., M. J. Ramírez Quintana, et al. (2007). Introducción a la Minería de Datos. España, Pearson Educación SA.
2. Ian H. Witten and Eibe Frank (2011). Data Mining: Practical Machine Learning Tools and Techniques, Third Edition. Morgan Kaufmann Series in Data Management Systems.
3. Jiawei Han and Micheline Kamber (2011). Data Mining: Concepts and Techniques, Third Edition. The Morgan Kaufmann Series in Data Management Systems.
4. Daniel T. Larose and Chantal D. Larose (2015). Data Mining and Predictive Analytics. Wiley Series on Methods and Applications in Data Mining.
5. Mohammed J. Zaki and Wagner Meira Jr (2014). Data Mining and Analysis: Fundamental Concepts and Algorithms. Cambridge University press.
6. Max Bramer (2013). Principles of Data Mining (Undergraduate Topics in Computer Science). 2nd Edition. Springer.
7. Pawel Cichosz (2015). Data Mining Algorithms. Wiley.
8. Pang-Ning Tan and Michael Steinbach (2005). Introduction to Data Mining. Pearson Editions.
9. David J. Hand and Heikki Mannila (2001). Principles of Data Mining (Adaptive Computation and Machine Learning) MIT Press.
10. Maimon, O. Z. and L. Rokach (2005). Data mining and knowledge discovery handbook. USA, Springer.
11. Pérez López, C. and D. Santín Gonzalez (2006). Data Mining-Soluciones Con Enterprise Miner. México, Alfaomega, Ra-Ma.
12. Sumathi, S. and S. N. Sivanandam (2006). Introduction to data mining and its applications. Berlín, Germany, Springer-Verlag New York Inc.
13. Everitt, B.S. (1994). A Handbook of Statistical Analyses using S-Plus. Chapman and Hall.
14. Kimball, R (1996). The Data Warehouse Toolkit. John Wiley & Sons.
15. Hastie, T., R. Tibshirani, et al. (2005). The elements of statistical learning: data mining, inference and prediction, Springer
16. Dunham. H. Margaret (2003). Data Mining. Introductory and Advanced Topics, Prentice Hall.
17. Pyle, D. (1999), "Data Preparation for Data Mining", Morgan Kaufmann, San Francisco, CA.