

1. Datos Generales de la asignatura

Nombre de la asignatura:	Ciencia de Datos
Clave de la asignatura:	CIC-2004
SATCA¹:	2-2-4
Carrera:	Ingeniería en Sistemas Computacionales

2. Presentación

Caracterización de la asignatura

Esta asignatura aporta al perfil del Ingeniero en Sistemas Computacionales, la capacidad para realizar aplicaciones basadas en la obtención de conocimiento a partir de bases de datos, mediante la utilización de técnicas dirigidas a la predicción y descripción de los mismos, con el propósito de integrar soluciones computacionales con diferentes tecnologías, plataformas y dispositivos; observándose siempre en el desempeño de sus actividades profesionales, los aspectos legales, éticos, sociales y de desarrollo sustentable.

El propósito del presente curso es el de complementar los conocimientos adquiridos en las materias de la estructura genérica sobre tratamiento de datos (Fundamentos de Base de Datos, Taller de base de datos y Administración de bases de datos), con la obtención de conocimiento mediante el uso de técnicas básicas de OLAP, así como de manejo de bases de datos NoSQL, Visualización de Datos y Minería de Datos.

Una característica importante de este curso es su orientación al uso de las tecnologías de Cloud Computing a fin de aprovechar los recursos disponibles, así como la creciente participación de las empresas en los servicios y soluciones albergadas en la nube.

Para cursar esta asignatura se requiere haber alcanzado competencias previas en SQL así como de desarrollo y programación de algoritmos relacionados con el tratamiento de datos. De igual manera el estudiante debe tener un conocimiento básico de idioma inglés a nivel técnico, a fin de comprender e implementar los procedimientos indicados en manuales específicos, para el uso de programas donde se incluyen las principales herramientas utilizadas en el Reconocimiento de Patrones.

Intención didáctica

A fin de obtener los resultados esperados, esta asignatura debe centrarse en dos vertientes, la primera se centra en el estudio del Data Warehouse e implementación de datamarts, mientras que, la segunda está dirigida al análisis y aplicación de algoritmos para la preparación, transformación y carga de datos para la posterior aplicación de técnicas descriptivas y predictivas, para lo que se requiere la realización de diversas prácticas, así como, de un proyecto del curso que pueda resolver un problema en

específico.

Un segundo resultado esperado es preparar al estudiante para cursar estudios avanzados y/o de posgrado, como especialidades o maestrías en el contexto de las ciencias computacionales.

Se organiza el temario, en cuatro unidades. La primera de ellas aborda diferentes aspectos a considerar para trabajar con almacenes de datos (Data warehouse) y herramientas OLAP, esto sirve como preámbulo para los siguientes tres temas que tratan de fundamentos de Minería de Datos (Data Mining) y que estarán centrados en la aplicación de métodos tanto supervisados como no supervisados. Asimismo, que se trabaje haciendo énfasis en técnicas de reducción de dimensionalidad para conjuntos de datos grandes, mediante selección de atributos y selección de instancias.

Durante todo el curso se realiza una fuerte inclinación para utilizar herramientas disponibles de diversas marcas o bien de tipo Open Source, aunque no se deja de lado que el desarrollo resultante de los proyectos pueda subirse a alguna nube como modelo de trabajo, para el ejercicio profesional en las empresas.

El enfoque sugerido para esta asignatura, requiere que las actividades prácticas promuevan el desarrollo de habilidades para el análisis de problemas que requieran el tratamiento de datos y la obtención del conocimiento implícito en los mismos. Así, en el último tema se pretende determinar la factibilidad de uso de recursos y /o servicios de la nube, para implementar un proyecto de Minería de Datos.

En el desarrollo de la materia, deberá observarse qué:

- Los contenidos sean abordados en su totalidad, procurando siempre que los alumnos cuenten con el material desarrollado por el docente (objetos de aprendizaje), de forma que puedan realizar trabajo fuera del laboratorio.
- El laboratorio de prácticas cuente con al menos dos SGBD que deberán utilizarse durante el desarrollo de la materia, así como el software WEKA para el procesamiento de las vistas minables.
- Toda práctica diseñada por el docente sea tomada con base al contexto de negocios de la región donde puede aplicarse el conocimiento adquirido.
- Los estudiantes sean capaces de utilizar estrategias de aprendizaje autodirigido, a fin de desarrollar el sentido de competitividad requerido en un entorno de productividad real.

3. Participantes en el diseño y seguimiento curricular del programa

Lugar y fecha de elaboración o revisión	Participantes	Observaciones
Instituto Tecnológico Superior del Oriente del Estado de Hidalgo, 24 de Enero de 2020.	Integrantes de la Academia del P.E. de Ingeniería en Sistemas Computacionales: <ul style="list-style-type: none"> • Dr. Lauro Vargas Ruiz • Dr. Elías Ruiz Hernández 	Definición de los programas de estudio de la especialidad de Computación en la Nube

4. Competencia(s) a desarrollar

Competencia específica de la asignatura
Procesar conjuntos de datos disponibles para obtener información y conocimiento útil mediante técnicas de OLAP y de Ciencia de datos con una actitud ética y colaborativa.

5. Competencias previas

<ul style="list-style-type: none"> • Crear y manipular 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. • Seleccionar y decidir la plataforma sobre la cual trabajará el sistema gestor de bases de datos. • Conocer, analizar e interpretar la filosofía de diseño de los sistemas operativos actuales y proponer aplicaciones programables para el manejo de los recursos que permitan facilitar el uso de la computadora a los usuarios potenciales.

6. Temario

No.	Temas	Subtemas
1	Tecnología OLAP para la Ciencia de Datos	1.1. Definición y conceptos relacionados. 1.2. Concepto de ETL, ejemplos a partir de distintas fuentes de datos. 1.3. Función y diseño de almacenes de datos y Datamarts. 1.3.1. Modelo estrella. 1.3.2. Copo de nieve. 1.4. Creación de Cubos de información. 1.5. Creación de Reportes (dashboards). 1.6. Uso de herramientas para Visualización de datos.
2	Fundamentos de Ciencia de Datos	2.1. Conceptos de Minería de Datos y KDD 2.2. Preprocesamiento de datos 2.3. Técnicas para discretización y transformación de datos. 2.4. Manejo de datos desconocidos o con ruido 2.5. Conceptos de Reconocimiento de patrones 2.5.1. Selección de variables 2.5.1.1. Métodos univariados y multivariados 2.5.2. Clasificación supervisada 2.5.3. Clasificación no supervisada
3	Métodos supervisados, no supervisados y de selección de atributos en Ciencia de Datos	3.1. Técnicas de aprendizaje Supervisado 3.1.1. Técnicas basadas en árboles (ID3, J48) 3.1.2. Técnicas basadas en Redes Neuronales 3.1.3. Técnicas basadas en aprendizaje estadístico (Redes Bayesianas) 3.1.4. Técnicas de regresión 3.2. Técnicas de Aprendizaje No Supervisado 3.2.1. Algoritmo k-medias 3.2.2. Algoritmo de clustering jerárquico

		<p>3.3. Técnicas de Selección de Atributos</p> <p>3.3.1. Algoritmos básicos de Selección de atributos (Ganancia de Información, PCA)</p> <p>3.4. Técnicas de Selección de instancias</p> <p>3.4.1. Algoritmos básicos para la selección de instancias</p>
4	Proyecto de Ciencia de Datos	<p>4.1. Selección del caso de estudio</p> <p>4.2. Evaluación de factibilidad de incorporar servicios y/o recursos Cloud Computing</p> <p>4.3. Presentación de propuesta incorporando extracción, limpieza, representación y análisis de los datos para facilitar la toma de decisiones de la organización.</p> <p>4.4. Desarrollo del proyecto</p> <p>4.5. Validación de resultados y presentación de conclusiones.</p>

7. Actividades de aprendizaje de los temas

Tema 1. Tecnología OLAP para la Ciencia de Datos	
Competencias	Actividades de aprendizaje
<p>Especifica(s):</p> <p>Conoce y comprende diferentes estructuras de datos, así como el concepto de OLAP y cómo estructurar los datos para crear un cubo con un modelado ETL previo.</p> <p>Genéricas:</p> <p>Competencias instrumentales</p> <ul style="list-style-type: none"> Habilidad en el manejo de equipo de cómputo. Comunicación oral y escrita en su propia lengua Conocimiento de una segunda lengua Habilidades de gestión de información <p>Competencias interpersonales</p> <ul style="list-style-type: none"> Capacidad crítica y de autocrítica. Capacidad de análisis y síntesis. Capacidad de trabajar en equipo o la expresión de compromiso ético. Solución de problemas <p>Competencias sistémicas</p> <ul style="list-style-type: none"> Capacidad de aplicar los conocimientos en la práctica. Habilidad para buscar y analizar información proveniente de fuentes diversas Capacidad para diseñar proyectos 	<ul style="list-style-type: none"> Investigar conceptos básicos del Procesamiento Analítico en Línea y su clasificación para elaborar un cuadro sinóptico. Hacer un resumen del origen de la tecnología OLAP. Discutir en plenaria las diferencias entre OLAP y OLTP, así como ventajas y desventajas. Investigar las diversas herramientas que existen para preparar datos (ETL), para crear cubos y realizar consultas sobre ellos. Desarrollar un proyecto de un caso real donde se implementen diversos cubos que integren un Datamart. Desarrollar un modelo de integración de servicios para ETL (Visual Studio) Desarrollar un reporte con gráficas vía alguna herramienta de visualización de datos (PowerBI, Rstudio, RapidMiner, entre otras).

Tema 2. Fundamentos de Ciencia de Datos	
Competencias	Actividades de aprendizaje
<p>Especifica(s):</p> <p>Comprende los diferentes conceptos básicos de Minería de Datos, incluyendo las etapas del KDD.</p> <p>Conoce y maneja diversas herramientas de minería de datos.</p> <p>Genéricas:</p> <p>Competencias instrumentales</p> <ul style="list-style-type: none"> Habilidad en el manejo de equipo de cómputo. Comunicación oral y escrita en su propia lengua Conocimiento de una segunda lengua Habilidades de gestión de información <p>Competencias interpersonales</p> <ul style="list-style-type: none"> Capacidad crítica y de autocrítica. Capacidad de análisis y síntesis. Capacidad de trabajar en equipo o la expresión de compromiso ético. Solución de problemas <p>Competencias sistémicas</p> <ul style="list-style-type: none"> Capacidad de aplicar los conocimientos en la práctica. Habilidad para buscar y analizar información proveniente de fuentes diversas Capacidad para diseñar proyectos 	<ul style="list-style-type: none"> Investigar conceptos de Minería de Datos y sus diferentes técnicas para hacer un cuadro sinóptico. Discutir en plenaria las diferencias entre Dato, Información y Conocimiento, así como las diversas formas con que se puede extraer. Realizar un modelo tabular de datos que pueda ser utilizado en Minería de Datos. Hacer un resumen de herramientas existentes para Minería de Datos. Analizar un caso genérico propuesto por el docente donde se utilice minería de datos. Utilizar herramientas básicas de minería de datos (Weka, Rapid Miner, Rstudio, Python, etc.) para corregir errores en tablas de datos.
Tema 3. Métodos supervisados, no supervisados y de selección de atributos en Ciencia de Datos	
Competencias	Actividades de aprendizaje
<p>Especifica(s):</p> <p>Conoce métodos supervisados y no supervisados para resolver una parte de un problema de Minería de Datos mediante trabajo colaborativo.</p> <p>Desarrolla su capacidad para implementar una solución que mediante un método supervisado o no supervisado predice una clase o agrupa elementos similares de un conjunto de datos.</p> <p>Conoce el uso y sentido de realizar selección de atributos a fin de reducir la dimensionalidad de los datos.</p> <p>Genéricas:</p> <p>Competencias instrumentales</p>	<ul style="list-style-type: none"> Estudiar las principales características de diversos métodos descriptivos y predictivos para observar el tipo de problemas que resuelven. Discutir en plenaria las diferencias entre métodos descriptivos y predictivos. Diferenciar las fronteras entre las disciplinas: Minería de Datos, Inteligencia Artificial y Reconocimiento de Patrones, Aprendizaje Profundo, Aprendizaje Máquina y otros. Exponer a manera de seminarios, artículos de divulgación científica sobre la aplicación de métodos descriptivos y predictivos en la resolución de problemas. Exponer un artículo científico que hable acerca de algún método de Aprendizaje supervisado o no supervisado.

<ul style="list-style-type: none"> Habilidad en el manejo de equipo de cómputo. Comunicación oral y escrita en su propia lengua Conocimiento de una segunda lengua Habilidades de gestión de información <p>Competencias interpersonales</p> <ul style="list-style-type: none"> Capacidad crítica y de autocrítica. Capacidad de análisis y síntesis. Capacidad de trabajar en equipo o la expresión de compromiso ético. Solución de problemas <p>Competencias sistémicas</p> <ul style="list-style-type: none"> Capacidad de aplicar los conocimientos en la práctica. Habilidad para buscar y analizar información proveniente de fuentes diversas Capacidad para diseñar proyectos 	<ul style="list-style-type: none"> Explica de manera general el funcionamiento de algún método de clasificación supervisada o no supervisada. Explica de manera general el funcionamiento de un algún método de selección de atributos. Utiliza un modelo que combina selección de atributos y clasificación (supervisada o no supervisada) para predecir una clase reduciendo el ruido de los datos.
Tema 4. Proyecto de Ciencia de Datos	
Competencias	Actividades de aprendizaje
<p>Especifica(s):</p> <p>Implementar en un programa de computadora una fase o un proyecto completo de Minería de Datos mediante trabajo colaborativo e individual.</p> <p>Genéricas:</p> <p>Competencias instrumentales</p> <ul style="list-style-type: none"> Habilidad en el manejo de equipo de cómputo. Comunicación oral y escrita en su propia lengua Conocimiento de una segunda lengua Habilidades de gestión de información <p>Competencias interpersonales</p> <ul style="list-style-type: none"> Capacidad crítica y de autocrítica. Capacidad de análisis y síntesis. Capacidad de trabajar en equipo o la expresión de compromiso ético. Solución de problemas <p>Competencias sistémicas</p> <ul style="list-style-type: none"> Capacidad de aplicar los conocimientos en la práctica. Habilidad para buscar y analizar información proveniente de fuentes diversas Capacidad para diseñar proyectos 	<ul style="list-style-type: none"> Proponer tres alternativas de proyecto del curso que requieran el desarrollo de métodos descriptivos o predictivos para elegir en conjunto con el docente el que se desarrollará. Revisar en la documentación de diversas nubes, los recursos y/o servicios disponibles para determinar la factibilidad de incorporarlos en el proyecto del curso. Implementar el software que integrará los métodos descriptivos o predictivos que resuelven el planteamiento dado, éste puede incluir sólo una parte o todo un modelo de Minería de Datos. Diseñar las pruebas que deberá satisfacer la aplicación desarrollada y documentar el resultado de las mismas

8. Prácticas

- Trabajar con una herramienta donde realice las tareas de Extracción, Transformación y Carga.
- Diseñar e implementar un programa en Java, Rstudio, Python o Visual Studio para realizar las funciones de Extracción, Transformación y Carga de datos provenientes de diferentes plataformas (Ubicaciones de red, Sistemas Operativos y/o SGBD, tal como fuentes SQL o NoSQL).
- Crear cubos a partir de bases de datos con diseño OLAP existentes.
- Crear un modelo Tabular de datos a partir de distintas fuentes de información
- Implementar o explicar la implementación hecha por un tercero, acerca de un método supervisado y uno no supervisado.
- Implementar un modelo con reducción de atributos vía selección usando algún lenguaje de programación y un kit o sdk para apoyarse en la tarea (java-weka, scikit-learn python, caret lenguaje R).
- Implementar un modelo con selección de instancias a fin de reducir el número de instancias en los datos y mostrar su utilidad para el conjunto de datos seleccionado.
- Implementar un proyecto de Minería de Datos (o una parte del mismo) de acuerdo a algún planteamiento validado por el docente que utilice selección de atributos con clasificación supervisada o no supervisada.

9. Proyecto de asignatura

El objetivo del proyecto que se proponga por 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 y 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; esto implica planificar un proceso de intervención empresarial, social o comunitario y el diseño de un modelo entre otros aspectos, así como de acuerdo con el tipo de proyecto se incluyen 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-profesional, 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 (específicas y genéricas de la asignatura)

<p>La evaluación es continua y formativa, por lo que en el desarrollo de la asignatura se promueve el trabajo por equipo y colaborativo a través de actividades de continua reflexión crítica.</p> <ul style="list-style-type: none"> • Conoce y comprende diferentes estructuras de datos, así como el concepto de OLAP y cómo estructurar los datos para crear un cubo con un modelado ETL previo. • Comprende los diferentes conceptos básicos de Minería de Datos, incluyendo las etapas del KDD. Conoce y maneja diversas herramientas de minería de datos. • Conoce métodos supervisados y no supervisados para resolver una parte de un problema de Minería de Datos mediante trabajo colaborativo. Desarrolla su capacidad para implementar una solución que mediante un método supervisado o no supervisado predice una clase o agrupa elementos similares de un conjunto de datos. Conoce el uso y sentido de realizar selección de atributos a fin de reducir la dimensionalidad de los datos. • Implementar en un programa de computadora una fase o un proyecto completo de Minería de Datos mediante trabajo colaborativo e individual. 	<p>Genéricas:</p> <p>Competencias instrumentales</p> <ul style="list-style-type: none"> • Habilidad en el manejo de equipo de cómputo. • Comunicación oral y escrita en su propia lengua • Conocimiento de una segunda lengua • Habilidades de gestión de información <p>Competencias interpersonales</p> <ul style="list-style-type: none"> • Capacidad crítica y de autocrítica. • Capacidad de análisis y síntesis. • Capacidad de trabajar en equipo o la expresión de compromiso ético. • Solución de problemas <p>Competencias sistémicas</p> <ul style="list-style-type: none"> • Capacidad de aplicar los conocimientos en la práctica. • Habilidad para buscar y analizar información proveniente de fuentes diversas • Capacidad para diseñar proyectos
---	--

11. Fuentes de información

<ol style="list-style-type: none"> 1. Berson A; Smith S, Data Warehousing Data Mining OLAP (Data Warehousing/Data Management), Mc- Graw-hill , Hardcover, 1997. 2. Han, J., Kamber, M., Data Mining, Concepts and techniques, Morgan Kaufmann, United States of America, 2001. 3. Harjinder S, G., Praskash C. R., Data Warehousing "La integración de la información para la mejor toma de decisiones", México, D.F. Prentice Hall Hispanoamericana, 1996. 4. Hernández, J. Ramírez, M.J. Ferri, C. Introducción a la Minería de datos, Pearson – Prentice Hall, 2004. 5. Marqués de Sa, J., Pattern Recognition, Concepts, Methods and Applications, Springer Verlag, Oporto, Portugal, 2001. 6. Mirkin, B. Clustering for data mining, a data recovery approach, Chapman & Hall/CRC, London, UK, 2005. 7. Witten, I. Frank, E. Hall, M. Data Mining: Practical machine learning tools and techniques, Fourth Edition, Morgan Kaufmann. 1993. 8. Bishop, Cristopher M., Pattern Recognition and Machine Learning (Information Science and Statistics). Springer-Verlag, Berlin Heidelberg. 2006 9. Ian Goodfellow, Yoshua Bengio, Aaron Courville. Deep Learning. The MIT Press, Cambridge, Massachusetts London, England, 2016.
