

MATERIA
MINERÍA DE DATOS Y DATA WAREHOUSING
Optativa III



Nivel de formación	Maestría en Tecnologías de Información				
Área de formación	Optativa abierta	Orientación	Gestión estratégica de TI		
Modalidad	Presencial	Carga horaria	48 hrs.	Créditos	7

Objetivo General

Que el alumno conozca los fundamentos teóricos de gestión de datos a través del proceso de Data-Warehousing (DW) y de la teoría de Minería de Datos (DM) a partir de la comprensión de técnicas y uso de herramientas que permiten aplicar los algoritmos mas populares usados en los paradigmas de DW y DM.

Objetivo Particular

- Que el alumno comprenda el marco general de la gestión de datos a través del uso de estrategias de DataWarehousing.
- Que el alumno comprenda el paradigma de análisis de datos utilizando técnicas de regresión, agrupamiento y clasificación.

Competencia y sub-competencia a desarrollar

Competencia genérica

Identificar los principales procesos y estrategias para la gestión de datos en el marco de un DataWarehouse que permita el posterior análisis de datos a través de técnicas de Minería de Datos.

Competencias específicas

- a) Identificar los procesos y estrategias para la gestión de datos y análisis de Datos
- b) Analizar las capas de estado de gestión de datos en un DataWarehouse y procesos utilizados para el análisis de Datos en un contexto de Data Mining
- c) Comprender las estrategias para la gestión de datos y las técnicas de análisis de datos más populares
- d) Implementaciones en relación con la gestión de datos con bases de datos relacionarles y NoSQL, entender y utilizar tanto técnicas de análisis de datos como algoritmos de minería de datos.

Producto esperado del curso

En el tema de DataWarehouse - Desarrollar una pequeña base de datos relacional, realizar la exportación de datos a un archivo separado por comas, importar un conjunto de datos en formato XML a una base de datos Relacional, realizar el modelado de una tabla de hechos con sus respectivas dimensiones.

En el tema de Minería de Datos - Realizar un diagrama de dispersión utilizando algún conjunto de datos, generar un análisis de correlación de variables, aplicar la regresión lineal al conjunto de datos, realizar un análisis de agrupamiento jerárquico, aplicar el algoritmo de agrupamiento particiones K-means a un conjunto de datos, realizar un análisis de clasificación utilizando herramientas para la obtención de un árbol de decisión.

Producto Final del Curso - una presentación que integre el análisis conjunto de datos con la aplicación de las técnicas antes mencionadas.



Campo de aplicación profesional

Los conocimientos adquiridos en la materia son utilizables en el campo de la gestión de datos empresariales, y del análisis de datos para la extracción de conocimiento sobre conjuntos de datos, conocimiento útil como soporte para la toma de decisiones.



Logros esperados

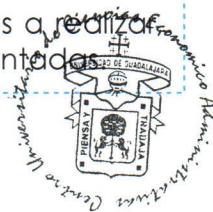
Conocimientos

El estudiante se apropiará de una serie de conceptos y principios que le permitirán identificar los procesos de un Data Warehouse y desarrollar un diseño básico de DW. Identificará las estrategias más populares utilizadas en el análisis de datos y la extracción de conocimiento de bases de datos.



Logros esperados

Habilidades	Será capaz de aplicar los conocimientos de data warehousing para la gestión de datos y del análisis de datos para la extracción de conocimiento de un conjunto de datos.
Actitudes	Disciplina y constancia en la aplicación de los conocimientos adquiridos. Inquietud y productividad en la investigación de conceptos y la manera de aplicarlos. Iniciativa en la búsqueda de resolución de problemas con relación a las actividades propuestas en clase.
Valores	Responsabilidad para con las actividades propuestas y tareas a realizar, respeto ante los presentes en el aula de clase y las ideas presentadas.



CONTENIDOS

Módulo / Unidad	Contenido	Producto o resultado esperado
Módulo I. Introducción a la ciencia de Datos	<ol style="list-style-type: none"> 1. Introducción a la ciencia de Datos 2. Descripción de la pirámide del conocimiento 3. Generalidades de Data Warehousing y Big Data 	Identificar la importancia de los Datos
Módulo II. Introducción a los DW	<ol style="list-style-type: none"> 1. Sistemas de Soporte para la toma de decisiones (Decision Support System - DSS) 2. Introducción a los conceptos de OLTP Y OLAP 3. Características y conceptos de un Data Warehouse 	Desarrollo de síntesis respecto de un DSS.
Módulo III. Arquitecturas de DW	<ol style="list-style-type: none"> 1. Arquitecturas de Data Warehouse (DW) 2. Propiedades de una arquitectura de DW 3. Descripción de arquitecturas de: Capa sencilla, doble capa, tres capas, y arquitecturas alternativas. 4. Procesos ETL y estructuras de datos: archivos planos, conjuntos de datos XML, Tablas relacionales, Metadatos. 5. Modelado Dimensional: Esquema tipo estrella, esquema de copo de nieve, Identificación de Dimensiones, Tabla de hechos. 	Desarrollo de una base de datos relacionar, capacidad para importar y exportar datos en tablas relacionales.

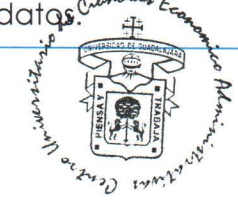


Módulo / Unidad	Contenido	Producto o resultado esperado
Módulo IV. Introducción al análisis de Datos - correlación y regresión	1. Introducción a la minería de datos 2. El proceso de minería de datos 3. Descripción del análisis de correlación entre variables. 4. Correlación Lineal 5. Diagramas de dispersión de datos 6. Coeficiente de correlación de Pearson 7. Análisis de regresión Lineal y ejemplo.	Desarrollo de un análisis de correlación de variables
Módulo V. Introducción a las estrategias de agrupamiento	1. Generalidades de los algoritmos de agrupamiento 2. Estrategias de agrupamiento jerárquico 3. Agrupamientos particionales: K-means 4. Distancia euclídea y no euclídea (Mahalanobis)	Desarrollo de un análisis de datos con estrategias de agrupamiento jerárquico y particional
Módulo VI. Clasificación de Datos	1. Introducción a la clasificación de datos 2. Características y diferencias entre agrupamiento y clasificación. Estrategias no supervisadas y estrategias supervisadas. 3. Bases de datos etiquetadas 4. Introducción a los árboles de decisión 5. Introducción a las Redes Neuronales	Aplicación de un algoritmo de árbol de decisión para obtener la función de clasificación de un conjunto de datos.
Módulo VII. Minería de Dato, soluciones para Big Data	1. Plataformas en línea para la gestión y análisis de Datos 2. Microsoft Azure ML	Identificar algunas plataformas que permiten la gestión de grandes volúmenes de datos.

BIBLIOGRAFÍA

Libros:

- **Golfarelli & Rizzi**, "Data Warehouse Design: Modern Principles and Methodologies", McGraw-Hill (2009), ISBN: 978-0-07-161039-1.
- **Ralph Kimball, Joe Caserta**, "The Data Warehouse ETL Toolkit", Wiley publishing Inc., (2004), eISBN: 0-764-57923-1.
- **Jared Dean**, "Big Data, Data Mining and Machine Learning: Value Creation for Business Leaders and Practitioners", (2014), Wiley & SAS Business Series, Published by John Wiley & Sons. ISBN 978-1-118-92069-5.



MAESTRÍA EN
TECNOLOGÍAS DE
INFORMACIÓN

Periférico Norte 799, Núcleo Universitario Los Belenes, Módulo "Q-301" Tercer Piso C.P. 45100
Zapopan, Jal., México. Tel: 01 (33) 37703440 Ext. 25326 Y 25327
Correo: mtiati@cucea.udg.mx posgrados.cucea.udg.mx

- **Cathy O’Neil, Rachel Schutt**, “*Doing Data Science: Straight Talk from the Frontline*”, (2013), O’Reilly Media. ISBN-13: 978-1449358655
- **Ian H. Witten, Eibe Frank, Mark A. Hall, Christopher J. Pal** “*Data Mining: Practical Machine Learning Tools and Techniques*”, Fourth Edition, (2017), Morgan Kaufmann, Elsevier. ISBN-13: 978-0128042915.

Criterios de Evaluación	Porcentaje / Puntaje
Productos de trabajo e investigación (Tareas)	30
Examen(es)	30
Proyecto Final	40
Total	100

Elaboró y actualizó el programa:
Última Revisión, actualización:

Dr. Leopoldo Gómez Barba
13 de julio del 2022

Aprobación de la Junta Académica
Programa de la Maestría en Tecnologías de Información



MAESTRÍA EN
TECNOLOGÍAS DE
INFORMACIÓN

Revisores:

Firma:

Dr. José Antonio Orizaga Trejo
Presidente de la Junta Académica

Dr. Sergio Roberto Dávalos García
Secretario

Dr. Víctor Manuel Larios Rosillo
Consejero

Dr. Cuauhtémoc López Martín
Consejero

Dra. María Elena Meda Campaña
Consejero

Mtro. Alejandro López Rodríguez
Consejero

